1990

# Failure time analyses for data collected from independent groups of correlated individuals

Sarah Margaret Nusser
*Iowa State University*

Order Number 9101373

# Failure time analyses for data collected from independent groups of correlated individuals

Nusser, Sarah Margaret, Ph.D.

Iowa State University, 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

# Failure time analyses for data collected from independent groups of correlated individuals

by

Sarah Margaret Nusser

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1990

## TABLE OF CONTENTS

# 1. INTRODUCTION

## 1.1. Introduction to the Problem

Failure time data are often collected from groups of correlated individuals. This can occur for example when data are collected from rat litters in teratology studies, from pairs of eyes in ophthalmology studies, or from members of families in medical studies. When individuals belong to such groups, independence among all observations cannot be validly assumed, and it is desirable to apply methods that account for the correlation among group members in the analysis of the failure times. Although failure time methodologies based on independent observations have been studied extensively (e.g., Kalbfleisch and Prentice, 1980; Lawless, 1980; Cox and Oakes, 1984), far fewer methods are available for analyzing correlated failure time data, and the methods that have been suggested are not as general as the independence-based analyses. Techniques that allow for explanatory variables are either restricted to small groups of responses, or they assume a very restricted intra-group correlation structure. Other methods that allow for larger group sizes and flexible correlation structures have not yet been extended to permit incorporation of explanatory variable information.

In the following chapters, a methodology is described that permits large and variable group sizes, heterogeneous intra-group correlations, and the use of explanatory variables. This method can be used with many types of failure time models, including proportional hazards models, and can be applied to data that are censored via a common interval censoring scheme (e.g., via a regular inspection schedule) or to exact time data, including possibly right censored data.

In this chapter, some basic principles of univariate failure time analyses are reviewed, and past research on multivariate survival methods for correlated data is summarized. The alternative methodology that forms the basis of this thesis is then briefly described.

## 1.2. Review of Univariate Failure Time Analysis Concepts

Failure time, survival or event time analyses are concerned with estimating the distribution of failure times (or time to the occurrence of an event) and/or determining the effects of explanatory variables on the failure time distribution. Failure time data typically involve some form of censoring due to, for example, subjects dropping out of a study, termination of an experiment before all individuals have failed, or inspection of individuals only at a specific finite set of time points.

Failure time analyses are generally modeled in terms of survival and hazard functions. Using T to denote the failure time, the survival function, $S(t)$, is defined to be

$$S(t) = \Pr\{ T \geq t \} .$$

The hazard function, $\lambda(t)$, describes the risk of failure in the near future given survival up to time t. For interval censored situations, where the entire time interval $[0, \infty)$ is divided into a set of disjoint intervals $\{[t_{h-1}, t_h) : h = 1, 2, \ldots, k+1; t_0 = 0; t_{k+1} = \infty\}$, $\lambda(t)$ is defined to be a step function whose steps are defined by

$$\lambda_h = \Pr\{ T \in [t_{h-1}, t_h) \mid T \geq t_{h-1} \} .$$

Often an underlying continuous failure time distribution with survival function $S(t)$ is used to express the discrete hazard as

$$\begin{aligned} \lambda_h &= [S(t_{h-1}) - S(t_h)] / S(t_{h-1}) \\ &= 1 - [S(t_h) / S(t_{h-1})] . \end{aligned}$$

Each step in the discrete hazard describes the conditional probability of failing in an interval given success up to the beginning of the interval. For a continuous failure time distribution with density $f(t)$, the hazard function describes the instantaneous probability of failure at time t given success up to time t. In this case,

$$\lambda(t) = \lim_{\Delta \to \infty} \Pr(\ T \in [t,\ t+\Delta)\ |\ T \geq t\ ) \ /\ \Delta$$

$$= f(t)\ /\ S(t)$$

$$= -\ \frac{\partial \log[S(t)]}{\partial t}\ .$$

Note that this definition implies that the survival function can be written as

$$S(t) = \exp\left\{ -\int_0^t \lambda(u)\ du \right\}.$$

Parametric distributions for failure time data analyses are typically defined for positive valued random variables. Some common examples include the gamma, Weibull, lognormal and log-logistic families. Explanatory variables can be included by modeling one (or more) of the distributional parameters as a function of the explanatory variables, or by making a proportional hazards assumption in which the hazard is assumed to be proportional to some function of the explanatory variables. Maximum likelihood is usually used to estimate the distributional parameters and the covariance matrix of the estimates.

Semi-parametric models are also available. Cox's proportional hazards model is frequently used when estimation of explanatory variable effects is the primary focus of the analyses. The hazard function is not completely specified, but is assumed to be proportional to some function of the explanatory variables. The most common form for the hazard function is

$$\lambda(t) = \lambda_0(t) \exp\{\underline{X}'\underline{\beta}\} ,$$

where $\lambda_0(t)$ is the unspecified baseline hazard function, $\underline{X}$ is the vector of explanatory variables, and $\underline{\beta}$ is the parameter vector associated with the explanatory variables. Partial likelihood techniques are usually used to estimate $\underline{\beta}$.

Nonparametric models for failure time distributions involve fitting a step function to the data. Kaplan-Meier (or product limit) estimation is often used to estimate the hazard function. Other techniques exist for obtaining estimates of the cumulative hazard function (see Cox and Oakes, 1984).

### 1.3. Multivariate Failure Time Analyses

Much of the literature for multivariate failure time analyses concentrate on the bivariate case, although several methods are extendible to larger group sizes. One of the earliest approaches to analyzing paired survival times is that of Holt and Prentice (1974). They extend Cox's proportional hazards model by placing the proportional hazards assumption on an unspecified baseline hazard for each pair. The correlation is thus treated as a nuisance parameter and emphasis is placed on estimating the explanatory variable effects. This method is limited in application to pairs of correlated observations.

Another area of development in multivariate failure time analyses involves random effects models. Clayton (1978) and Oakes (1982) suggest modeling the failure time distribution for each pair member with a proportional hazards assumption. The hazard for each member of a pair is assumed to be proportional to a function of an unobservable covariate which has a common value for both members of that pair. Conditional on the value of the random variable, the failure time results for the two members of any pair are independent. An unconditional bivariate distribution is obtained by averaging the product of the hazards for each member of the pair with respect to a gamma distribution assumed for the unobservable random variable. The resulting unconditional bivariate distribution depends on an association parameter that can also be interpreted as a relative risk. Clayton shows that the association parameter is equal to the unconditional hazard for the first member of the pair at time t given that the second member fails at time t, divided by the unconditional hazard for the first member at time t given that the second fails after time t;

it can be alternatively defined by exchanging the roles of the members of a pair in the definition. The model proposed by Clayton and Oakes is developed for parametric models, assumes a common correlation for each pair, and can only be applied to uncensored data.

Clayton and Cuzick (1985) extend the Clayton-Oakes distribution to include fixed explanatory variables in the proportionality function that links the unobservable random variable to the hazard for each group member (before averaging over the distribution of the unobservable random variable). However, the marginal hazard from the bivariate distribution (i.e., the distribution that results from averaging the product of the hazards for each member of the pair over the distribution of the unobservable random variable) does not follow a proportional hazards assumption. Hougaard (1986) notes that the explanatory variable parameters in this model are confounded with the association parameter, so that the association parameter is measuring more than dependence. Huster, Brookmeyer and Self (1989) also extend the Clayton-Oakes model to include explanatory variables, but do so by placing a proportional hazards assumption on the marginals of the parametric unconditional bivariate distribution. Their estimation procedures are considerably simpler than Clayton and Cuzick's, and their methods allow for censored data. Oakes (1989) extends the

Clayton-Oakes distribution to include negative correlations. All of these random effects models assume a constant intra-group correlation and are theoretically extendible to more than two members per group, although the tractability of the derivations for higher dimensions is not clear.

Hougaard (1986) follows a similar random effects approach using a positive stable distribution for the unobservable random variable. Although his model is restricted to positive correlations that are constant across groups, it is extendible to hierarchical intra-group correlations (e.g., when twins are more highly correlated than other siblings). In addition, Hougaard is able to use explanatory variables by placing a parametric or semi-parametric proportional hazards assumption on the marginals of the unconditional multivariate distribution. Parameters for the parametric proportional hazards model are estimated by maximizing the appropriate likelihood function. An alternative algorithm for estimating the explanatory variable parameters in a semi-parametric Cox proportional hazards model is suggested by Hougaard, although its statistical properties are unknown. Crowder (1989) develops a similar model based on Weibull assumptions that allows negative correlations, although it assumes homogeneous intra-group correlations. Both models are adaptable to groups with more than two members.

Recently, several researchers have explored an approach that uses a working model based on the incorrect assumption of completely independent responses (often referred to as an "independent working model"). Consistent estimates for the parameters of the marginal failure time distributions are obtained using methods that assume independence among all observations, but the covariance matrix for the parameter estimates is estimated using methods that allow for the presence of correlation among responses. The intra-group correlations are nuisance parameters under this approach, and thus can follow an arbitrary structure; however, inference on the association among group members is not possible.

Huster et al. (1989) and Wei and Amato (1989) both use robust estimation to obtain estimates of the covariance matrix for the parameter estimates. They develop covariance matrix estimators for parametric and semi-parametric proportional hazards models, respectively. Wei and Amato's derivations specifically assume that the group sizes are small relative to the number of groups. Huster et al. use simulations to assess the performance of this approach and find that it can be highly inefficient when intra-group correlations are high and/or if members of the pairs have the same explanatory variable values (e.g., received the same treatment).

Koehler and McGovern (1990) use bootstrap methods to estimate the covariance matrix of the parameter estimates by resampling the groups. This allows the groups to be of large and variable size, but their procedures have not been extended to include explanatory variables.

All of the methods above, except that of Koehler and McGovern (1990), are limited in application to small group sizes. In addition, excepting Hougaard's (1986) work, the random effects methods assume that correlations are homogeneous within and across groups. Independent working model methods allow for heterogeneous correlations, but correlations become nuisance parameters, excluding any investigations of association among group members. If the degree of correlation is not of interest, Koehler and McGovern's (1990) approach is appealing because it permits heterogeneous correlation structures and large group sizes, but it is limited by its inability to incorporate explanatory variables.

## 1.4. Proposed Methodology for Correlated Failure Time Data

The following chapters describe a method of estimating failure time distributions for data collected from independent groups of correlated individuals. The technique allows for large and variable group sizes, heterogeneous correlation structures, and the incorporation of explanatory variable information. Both parametric and nonparametric failure time models can be estimated, and correlations may be modeled as well.

The analyses rely on representing the failure times as conditional binary variables indicating the failure of an individual during a specified interval given success in the previous interval. For each time interval, a vector of binary responses is constructed for each group, consisting of responses for individuals belonging to the group who are at risk at the beginning of the interval and not censored during the interval. Each vector of responses has a mean vector whose elements are hazard probabilities and are thus functions of the failure time distribution parameters. The covariance matrix for a vector of binary responses is a function of the corresponding mean vector and parameters describing the correlations among the elements of the observed response vector. To obtain estimates of the failure time distribution parameters, multivariate nonlinear least squares estimation is used based on a Gauss-Newton algorithm. When the Gauss-Newton iterations are initiated with consistent estimates of the mean model (i.e., failure time distribution) and correlation parameters, the estimated mean model parameters have a joint asymptotic normal distribution under mild regularity conditions. Furthermore, since this estimation procedure uses information on covariances, it may be more efficient than estimators based on a working model that assumes independent responses.

Chapters 2 and 3 describe the methodology in detail for the common interval censoring and exact time cases. Examples of applications are given for both types of data. Asymptotic properties of the estimators are developed in Chapter 4. In Chapter 5, the properties and performance of several estimators of correlation coefficients for clustered binary data are discussed.

# 2. FAILURE TIME ANALYSES FOR INDEPENDENT GROUPS OF CORRELATED INDIVIDUALS UNDER A COMMON INTERVAL CENSORING SCHEME

## 2.1. Introduction

Survival studies based on groups of individuals generally yield correlated outcomes within groups. Correlations may arise from social relationships among members of a group, such as pairs of spouses, or from genetic relationships, such as groups of siblings. Data with this structure may involve groups of varying sizes, and correlations within groups may not be identical for all pairs. Interval censoring occurs when individuals are checked at the end of specific time intervals to determine whether the individual has failed since the previous inspection time. In such cases, the exact time to censoring or failure is not observed, but the event is known to have occurred within a particular interval.

This chapter introduces a failure time methodology for commonly interval censored data collected from independent groups of correlated individuals. The approach allows for variable group sizes and heterogeneous correlations among individuals within groups. In addition, a wide range of failure time models may be used, including models that depend on explanatory variables.

In this approach, the outcome for any individual at risk during a specific interval is modeled as a conditional binary variable indicating the failure or success of the individual given success in the previous interval. For each time interval, a separate vector of binary responses is constructed for each group, whose length is equal to the number of individuals in the group who are at risk or are not censored during the interval. The elements of the corresponding mean vector are hazard probabilities and are thus functions of the failure time distribution parameters. The covariance matrix for each binary response vector is a function of the mean vector and parameters describing the correlation structure among the elements of the observed response vector. The parameters of the failure time distribution are estimated using least squares estimation for multivariate nonlinear models.

This chapter describes how the binary response vector and associated mean vector and covariance matrix for each group and interval are constructed. Estimators for the failure time distribution parameters are presented, followed by a discussion of specific models for the hazard probabilities. The approach is then illustrated with an analysis of data from a study in which three smoking cessation programs are compared.

## 2.2. Binary Response Vector Definition and Distribution

Suppose that the data consist of observations from m independent groups of individuals and that failure or censor times for each individual are observed to fall into one of $k+1 < \infty$ disjoint intervals, $\{[t_{h-1}, t_h) : h = 1, 2, \ldots, k+1; t_0 = 0; t_{k+1} = \infty\}$, where time $t_k$ represents the last time of inspection or follow-up. Let $n_{hi}$ be the number of individuals in the risk set for group i at time $t_{h-1}$ minus the number of individuals who are censored during interval h, and let $m_h$ denote the number of groups in interval h with $n_{hi} > 0$. Note that $n_{k+1,i} = 0$ for all i and $m_{k+1} = 0$. Define $Y_{hij}$ such that

$$
\begin{aligned}
Y_{hij} &= 1 \quad \text{if individual j in group i fails during interval h given success up to } t_{h-1}, \\
&= 0 \quad \text{if individual j in group i succeeds during interval h given success up to } t_{h-1},
\end{aligned}
$$

where $i = 1, 2, \ldots, m_h$ groups and $j = 1, 2, \ldots, n_{hi}$ individuals. $Y_{hij}$ is not defined for any individual who is censored during interval h or has failed or been censored before $t_{h-1}$. This implies that $Y_{k+1,ij}$ is undefined for all i and j, so that the sums over h in the following sections end at $h = k$.

The mean and variance for $Y_{hij}$ can be constructed by observing that $Y_{hij}$ is a Bernoulli random variable with a mean equal to the hazard probability for individual j in group i during interval h. The hazard probability is derived from assumptions on the failure time distribution. Using $T_{ij}$ to represent the continuous exact failure time for individual j in group i with hazard function $\lambda_{ij}(t)$ and survival function $S_{ij}(t)$,

$$\begin{aligned} E(Y_{hij}) &\equiv \pi_{hij} \\ &= \Pr\{\text{individual } j \text{ in group } i \text{ fails in interval } h \text{ given success up to } t_{h-1}\} \\ &= \Pr\{\, T_{ij} \in [t_{h-1}, t_h) \mid T_{ij} \geq t_{h-1} \,\} \\ &= 1 - [S_{ij}(t_h) \,/\, S_{ij}(t_{h-1})] \\ &= 1 - \exp\left\{ - \int_{t_{h-1}}^{t_h} \lambda_{ij}(s)\, ds \right\} . \qquad (2.1) \end{aligned}$$

The variance of $Y_{hij}$ is $\pi_{hij}(1-\pi_{hij})$.

The binary variables can be used to construct an $n_{hi} \times 1$ response vector,

$$\underset{\sim}{Y}_{hi} = (Y_{hi1}, Y_{hi2}, \ldots, Y_{hin_{hi}})' ,$$

for each group and each interval for which $n_{hi} > 0$. The mean vector for $\underset{\sim}{Y}_{hi}$ is

$$\underline{\pi}_{hi} = (\pi_{hi1}, \pi_{hi2}, \ldots, \pi_{hin_{hi}})' \ .$$

The covariance matrix for $\underline{Y}_{hi}$, denoted $V_{hi}$, consists of variances $\pi_{hij}(1-\pi_{hij})$ along the diagonal and covariances

$$\rho_{hijj'} \ [\pi_{hij}(1-\pi_{hij})\pi_{hij'}(1-\pi_{hij'})]^{1/2} \ ,$$

on the off-diagonals, where $\rho_{hijj'}$ is the correlation between $Y_{hij}$ and $Y_{hij'}$. Because groups are assumed to be independent, individuals in different groups ($i \neq i'$) have a covariance of zero. Also, by conditioning on previous responses, observations from different intervals have zero covariance.

## 2.3. Parameter Estimation

### 2.3.1 Model for $\underline{Y}_{hi}$

The observed response vector can be modeled as a function of its mean vector plus a vector of errors. In general, the mean vector $\underline{\pi}_{hi}$ is a nonlinear function of the parameters, $\underline{\gamma}$, and possibly explanatory variables, $\underline{X}_{hi}$, that define the failure time distribution. In addition, the covariance matrix for the response vector is a function of the mean vector and another set of parameters, $\underline{\alpha}$, associated with the correlation coefficients for individuals in the response vector. For the purposes of the model definition, let $\pi_{hij} = \pi(\underline{\gamma}, \underline{X}_{hij})$, $\underline{\pi}_{hi} = \underline{\pi}(\underline{\gamma}, \underline{X}_{hi})$, and $V_{hi} = V(\underline{\theta}, \underline{X}_{hi})$, where $\underline{\theta} = (\underline{\alpha}', \underline{\gamma}')'$.

The model for $\underline{Y}_{hi}$ is assumed to be

$$\underline{Y}_{hi} = \underline{\pi}(\underline{\gamma}, \underline{X}_{hi}) + \underline{e}_{hi} , \qquad (2.2)$$

where $\underline{\gamma}$ is an s x 1 vector of fixed, unknown parameters belonging to the parameter space $\Gamma$; $\Gamma$ is a compact subset of $\mathbb{R}^s$; $\underline{X}_{hi} = (\underline{X}'_{hi1}, \ldots, \underline{X}'_{hin_{hi}})'$ is an $rn_{hi}$ x 1 vector belonging to a compact subset of $\mathbb{R}^{rn_{hi}}$ and containing the r x 1 explanatory variable vectors associated with each of the $n_{hi}$ individuals contributing to the hi-th observed response vector; $\underline{\pi}$ is an $n_{hi}$ x 1 vector whose elements are continuous functions from $\Gamma \times \mathbb{R}^{rn_{hi}}$ into [0,1] with continuous and uniformly bounded first and second and continuous third partial derivatives with respect to $\underline{\gamma}$; the $\underline{e}_{hi}$ are independent across h and i with mean $\underline{0}$ and nonsingular covariance matrix $V(\underline{\theta}, \underline{X}_{hi})$ for $\underline{\theta} = (\underline{\gamma}', \underline{\alpha}')$; $\underline{\alpha}$ is a u x 1 vector of fixed, unknown parameters belonging to parameter space $\Phi$; and $\Phi$ is a compact subset of $\mathbb{R}^u$.

### 2.3.2. Estimating the Mean Model Parameters

If the covariance matrices for $\underline{Y}_{hi}$ are known, multivariate nonlinear least squares estimation can be used to obtain estimates of the parameters in $\underline{\gamma}$. This is achieved by minimizing the weighted residual sum of squares

$$Q(\underline{\gamma}) = k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} \{ [\underline{Y}_{hi} - \underline{\pi}(\underline{\gamma}, \underline{X}_{hi})]' \, V(\underline{\theta}, \underline{X}_{hi})^{-1}$$

$$\times [\underline{Y}_{hi} - \underline{\pi}(\underline{\gamma}, \underline{X}_{hi})] \}$$

with respect to $\underline{\gamma}$.

In the usual case where the $V_{hi}$ are unknown, a Gauss-Newton algorithm can be used to obtain nonlinear least squares parameter estimates. The algorithm is derived from a first order approximation to $\pi_{hij}$ (see Chapter 4). The estimator for $\underline{\gamma}$ is calculated using an iterative process in which the c-th step is defined by

$$\hat{\underline{\gamma}}^{(c)} = \hat{\underline{\gamma}}^{(c-1)}$$

$$+ W^{-1}(\hat{\underline{\theta}}^{(c-1)}) \; k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} \{D(\hat{\underline{\gamma}}^{(c-1)}, \underline{X}_{hi})'$$

$$\times V^{-1}(\hat{\underline{\theta}}^{(c-1)}, \underline{X}_{hi}) [\underline{Y}_{hi} - \underline{\pi}(\hat{\underline{\gamma}}^{(c-1)}, \underline{X}_{hi})]\},$$

where

$$D(\underline{\gamma}, \underline{X}) = \partial \underline{\pi}(\underline{\gamma}, \underline{X}) / \partial \underline{\gamma}' \; ,$$

$$W(\underline{\theta}) = k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} D(\underline{\gamma}, \underline{X}_{hi})' \; V^{-1}(\underline{\theta}, \underline{X}_{hi}) \; D(\underline{\gamma}, \underline{X}_{hi}) \; ,$$

$$\hat{\underline{\theta}}^{(c-1)} = (\hat{\underline{\gamma}}^{(c-1)\prime}, \hat{\underline{\alpha}}^{(0)\prime})' ,$$

and $\hat{\underline{\alpha}}^{(0)}$ is the initial value for $\underline{\alpha}$ (which could alternatively be updated at each iteration). By initiating the Gauss-Newton iterations with a consistent estimator of

$\underline{\theta}$, say $\hat{\underline{\theta}}^{(0)} = (\hat{\underline{\gamma}}^{(0)\prime}, \hat{\underline{\alpha}}^{(0)\prime})'$ (where consistency is achieved as the number of groups increases), the conditions cited for model (2.2) plus some mild regularity conditions are sufficient to show that the Gauss-Newton estimator terminated at any step c is asymptotically normal with mean $\underline{\gamma}$ and covariance matrix $[\text{km } W(\underline{\theta}_0)]^{-1}$ (see Chapter 4). A consistent estimator for the covariance matrix of $\underline{\gamma}$ is $(\text{km } \hat{W})^{-1}$, where

$$\hat{W} = k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} D(\hat{\underline{\gamma}}, \underline{X}_{hi})' \; V^{-1}(\hat{\underline{\theta}}, \underline{X}_{hi}) \; D(\hat{\underline{\gamma}}, \underline{X}_{hi}) \; ,$$

$\hat{\underline{\theta}} = (\hat{\underline{\gamma}}', \hat{\underline{\alpha}}^{(0)\prime})'$, and $\hat{\underline{\gamma}}$ is the least squares estimator of $\underline{\gamma}$ obtained from the Gauss-Newton algorithm.

### 2.3.3. Consistent Initial Estimators for the Mean Model Parameters

Consistent initial estimators for the parameter vector in the mean model, $\hat{\underline{\gamma}}^{(0)}$, can be obtained by either of two methods. First, standard estimation procedures based on the working assumption of independent observations for the assumed survival model (e.g., maximum likelihood, partial likelihood or nonparametric methods) can be used to obtain estimates of the failure time distribution parameters from the original failure time data. Although these methods assume independence, they are generally consistent when data are correlated. For example, Huster, Brookmeyer and Self

(1989) and Wei and Amato (1989) demonstrate the consistency under an arbitrary correlation model of the parameter estimators for parametric and semi-parametric proportional hazards models using maximum and partial likelihood methods, respectively.

Alternatively, nonlinear regression techniques can be used to provide consistent initial estimators for $\boldsymbol{\gamma}$ by regressing the individual binary responses $Y_{hij}$ on the appropriate mean model. Dummy variables are used to incorporate the mean models for the binary responses from all time intervals into a single model as follows. Let

$$Z_{hij} = \begin{cases} 1 & \text{if data arise from individual } j \text{ in group } i \\ & \text{in interval } h \text{ (i.e., current dependent} \\ & \text{variable is } Y_{hij}) \\ 0 & \text{otherwise ,} \end{cases}$$

where $h = 1, 2, \ldots, k$ , $i = 1, 2, \ldots m_h$ , and $j = 1, 2, \ldots n_{hij}$ . Then

$$Y_{hij} = Z_{1ij}\pi_{1ij} + \cdots + Z_{kij}\pi_{kij} + e^*_{hij} ,$$

where $e^*_{hij}$ is an error term with zero mean. The entire binary data set can be used in a nonlinear regression program that assumes independence among the observations to obtain a consistent estimate of $\boldsymbol{\gamma}$. The nonlinear regression approach may be computationally expensive for large data sets or complicated mean models. In these cases, the first approach may be more easily and successfully implemented.

### 2.3.4. Consistent Initial Estimators for the Correlation Parameters

Auxiliary parameters in the covariance matrix consist of correlation coefficients or parameters from a model for the correlation coefficients. Consistent estimators of correlation coefficients are developed in Chapter 5 for clustered binary data with constant intra-group variances and for arbitrary intra-group variance structures. Recall that consistency is associated with increasing the number of groups. Hence, these estimators are appropriate when models for intra-group correlations can be applied within reasonably large subsets of the groups in the sample.

When the common correlation model is inadequate, it may be possible to partition all possible intra-group pairs of individuals into distinct classes, each with a distinct common correlation coefficient. For example, classes may correspond to the sexes of pairs of litter mates (female-female, male-male, female-male), or to social and biological relationships among human subjects. The consistency of the correlation estimators for different correlation classes is based on pooling information from intra-group pairs of individuals belonging to the same correlation class in a number of different groups. Under these conditions, the estimators of correlation developed in Chapter 5 can be applied to each correlation class according to the assumptions under which the specific estimator is consistent.

The correlation coefficient may also be modeled as a function of explanatory variables. For example, one possible model is

$$2^{-1}(\rho_{hijj'} + 1) = [1 + \exp\{\underline{Z}'_{hijj'}\underline{\alpha}\}]^{-1} \exp\{\underline{Z}'_{hijj'}\underline{\alpha}\} \ , \tag{2.3}$$

where $\underline{Z}_{hijj'}$ is a vector of explanatory variables associated with pair jj′ in group i during interval h, and $\underline{\alpha}$ is the corresponding vector of parameters. Explanatory variables may include continuous and classification variables, and possibly functions of time or order. The parameters $\underline{\alpha}$ in model (2.3) may be estimated using a nonlinear regression program with data generated by substituting

$$\hat{\rho}_{hijj'} = \frac{(Y_{hij} - \hat{\pi}_{hij})(Y_{hij'} - \hat{\pi}_{hij'})}{[(Y_{hij} - \hat{\pi}_{hij})^2(Y_{hij'} - \hat{\pi}_{hij'})^2]^{1/2}}$$

$$= \frac{(Y_{hij} - \hat{\pi}_{hij})(Y_{hij'} - \hat{\pi}_{hij'})}{|(Y_{hij} - \hat{\pi}_{hij})(Y_{hij'} - \hat{\pi}_{hij'})|} \ ,$$

or some other suitable estimator for $\rho_{hijj'}$ into the left hand side of equation (2.3). Although the estimator for $\underline{\alpha}$

is evaluated under the incorrect assumption that all estimated pairwise correlation coefficients are independent, it still produces a consistent estimate of $\underline{\alpha}$.

Because of requirements necessary to achieve consistency (i.e., averaging over groups), it is not possible to form consistent estimators for models where each group has its own arbitrary level of correlation among pairs of individuals. If inconsistent estimators for the correlation coefficient are used, then the mean model parameter estimators may not retain the asymptotic properties derived in Chapter 4.

## 2.4. Examples of Mean Models

### 2.4.1. Models Without Explanatory Variables

Either parametric or nonparametric models can be used to model $\pi_{hij}$ in the absence of explanatory variables. A simple nonparametric model can be developed by defining the hazard probability for interval h to be an arbitrary constant $\lambda_h$. The vector of hazard probabilities for this case is modeled as

$$\pi_{hij}(\underline{\gamma}) = \lambda_h ,$$

where $\underline{\gamma} = (\lambda_1, \lambda_2, \ldots \lambda_k)$. Alternatively, the mean can be constructed by assuming a parametric distribution and using equation (2.1). Two-parameter distributions are commonly used for this purpose, including the Weibull, gamma and lognormal distributions.

Both the nonparametric and parametric models assume a homogeneous failure time curve for all individuals, but can be easily modified to allow for different failure time distributions for distinct classes of individuals by using a distinct set of parameters for each class.

### 2.4.2. Proportional Hazards Models

It is often of interest to include explanatory variables in the analysis of failure time data. One common regression model is the proportional hazards model. An assumption of proportional hazards implies that each individual's hazard function $\lambda(t)$ is proportional to a baseline hazard function $\lambda_o(t)$. The proportionality constant is typically taken to be $\exp\{\underline{X}'\underline{\beta}\}$, so that the hazard function for an individual with r explanatory variables $\underline{X}$ is

$$\lambda(t) = \lambda_o(t) \exp\{\underline{X}'\underline{\beta}\} .$$

In the interval censored case, an alternative expression for the baseline survival function $S_o(t)$ is useful in constructing the proportional hazards model. $S_o(t)$ can be written as a function of a discrete baseline step function $\{\lambda_{oh} : h = 1, 2, \ldots, k\}$, whose steps correspond to the hazard probability for each interval:

$$S_o(t) = \prod_{h:t_{h-1}<t} [1 - P_o\{T \in [t_{h-1}, t_h) | T \in [t_{h-1}, \infty)\}]$$

$$= \prod_{h:t_{h-1}<t} (1 - \lambda_{oh})$$

(Kalbfleisch and Prentice, 1980). Thus, under the proportional hazards assumption, the survival function $S(t,\underline{X})$ is expressed as

$$S(t,\underline{X}) = S_o(t)^{\exp\{\underline{X}'\underline{\beta}\}}$$

$$= \prod_{h:t_h \geq t} (1 - \lambda_{oh})^{\exp\{\underline{X}'\underline{\beta}\}} ,$$

implying that the mean model is

$$\pi(\underline{\gamma},\underline{X}_{hij}) = 1 - (1 - \lambda_{oh})^{\exp\{\underline{X}_{hij}'\underline{\beta}\}} .$$

where $\underline{\gamma} = (\lambda_{o1}, \lambda_{o2}, \ldots, \lambda_{ok}, \underline{\beta}')'$. As noted in Section 2.4.1, the discrete baseline hazard may be assumed to be an arbitrary step function or it can be derived from a parametric assumption on the baseline failure time distribution. The semi-parametric Cox proportional hazards model cannot be applied in this framework because the unspecified baseline hazard does not provide enough information to develop an explicit function for the mean model.

### 2.4.3. Other Regression Models

It is possible to construct other regression models by placing models on the parameter(s) of any parametric distribution. One common model is the accelerated failure time model, which assumes that failure times follow a Weibull distribution with a log-linear model on the scale parameter. The Weibull accelerated failure time model is a proportional hazards model where the baseline hazard $\lambda_o(t)$ is assumed to follow a Weibull distribution; however, this relationship between the accelerated failure time model and proportional hazards model does not exist for most other parametric assumptions.

### 2.4.4. Time-Dependent Explanatory Variables

By viewing the hazard probabilities as conditional on the past process of stochastic variables and success up to the beginning of the interval, both internal and external time-dependent explanatory variables (sensu Kalbfleisch and Prentice, 1980) can be incorporated into the mean model, and the associated parameter estimates can be obtained via multivariate nonlinear least squares estimation as described above. For internal time-dependent variables, however, conditioning on the past process of the variable implies that the usual relationship between the hazard function and the survival function no longer exists.

## 2.5. Application of Methodology to Smoking Cessation Data

### 2.5.1. Overview

Treatment programs to help individuals quit smoking can be evaluated by collecting information on recidivism from participants in the program. Such data are often analyzed using categorical methods on the observed frequency of recidivists and abstainers. If actual dates of recidivism or information on smoking status at various points in time are known, the success of smoking cessation programs can be assessed using failure time methods. For smoking cessation data, time zero represents the initial quit date and the failure event is defined to be the resumption of the smoking habit according to some criterion of recidivism.

Data from smoking cessation studies arise from a mixture of two subpopulations: failers (recidivists) and permanent abstainers. The relative size of these two subpopulations in the treatment population is one measure of the success of a smoking cessation treatment. If all of the potential failures in a sample are observed, then the estimated relative size of the failing subpopulation is the observed proportion of recidivists. However, it is rarely known whether all failures have occurred during the study period, and if the recidivism data for an individual is right censored, it is not known to which subpopulation the individual belongs. Hence it is necessary to use more sophisticated techniques to estimate the proportion of the failing subpopulation.

In addition to the success rate of a treatment program as measured by the proportion of failers, failure time analyses provide valuable information on the patterns of recidivism over time. For example, it may be determined that one program has an extremely high recidivism rate at a particular phase in the treatment process, indicating the need for more intensive intervention measures during that phase of the program. Failure time analyses that incorporate explanatory variable information also provide a convenient means for determining the effects of various factors on the shape of the recidivism hazard function.

In the sections that follow, the failure time methods outlined in this chapter are applied to data from a comparative evaluation of smoking cessation clinics. The study is described, and appropriate mean models are developed. Results of the failure time analyses are then presented and discussed.

### 2.5.2. Description of Smoking Cessation Clinic Study

The effectiveness of three smoking cessation programs was evaluated in a study conducted in Iowa. This study is described in detail by Lando, McGovern, Barrios and Etringer

(1990). The clinics under consideration were developed by the American Cancer Society (ACS), the American Lung Association (ALA), and by Dr. Harry Lando (LANDO) (Lando, 1977; Lando and McGovern, 1982). All three programs were conducted in Des Moines, Iowa City and Waterloo. Treatment programs were administered to groups of individuals. The groups provided a framework in which participants discussed the benefits of abstinence, coping strategies for quitting, and maintaining a smoke-free life style. The format for addressing these issues varied with smoking cessation treatment.

The ACS program consisted of an orientation session plus four one-hour sessions over a two week period. Each facilitator was responsible for developing a clinic format that fostered active involvement of group members and that addressed the individual members' needs. No target date was set for quitting. Although participants were expected to quit in the latter half of the program, they were informed that smokers who take two weeks to quit are as likely to be successful as those who take two months to quit.

The ALA clinics involved an orientation session plus seven one-and-a-half to two-hour sessions over a seven week period, with a target quit date set at the third session. The format of the program was specifically outlined by the ALA. The first four sessions covered specific topics related to the quitting process (session 1: health effects; session 2: coping strategies; session 3: declaration of commitment to quitting on the actual quit date; session 4: reiteration of quitting benefits and discussion of withdrawal symptoms). Meetings 5 and 6 focused upon maintenance and development of healthy enjoyable nonsmoking lifestyles. A celebration was planned for the seventh meeting to reward the successful participants.

The LANDO program consisted of sixteen three-quarter to one-hour sessions over nine weeks, again with a relatively specific agenda for the meetings. The first eight sessions were held in the first three weeks, and focused upon preparation for quitting. The treatment included nicotine fading by use of increasingly strong cigarette filters or by switching to lower nicotine brands as the quit date approached. After the quit date, the second eight sessions were conducted over the remaining six weeks to help participants maintain abstinence. Group sessions consisted of relatively unstructured group discussion with emphasis on problem solving. Participants also signed contracts calling for specific rewards for abstinence.

Data on recidivism were collected during the treatment program. After the treatment program was terminated, follow-up contacts were made for each participant at 3, 6, 9, 12, 18, 24 and 36 months after the quit date. If at the time of contact, the participant admitted to having smoked at least one puff per day for seven consecutive days since the last follow-up contact, the participant was considered to have failed the quit attempt during the preceding time interval. If participants could not be reached, the data were considered to be right censored. Individuals who did not abstain initially for at least 24 hours were not considered to have quit and were omitted from the analyses.

Recidivism dates were often reported by the participant with some degree of imprecision. For example, an individual who resumed smoking at "three weeks" may have resumed anywhere between two-and-a-half and three-and-a-half weeks. To account for this, the following intervals for reported failure or censor times in units of days were constructed: first few days [0, 4), one week [4, 11), two weeks [11, 18), three weeks [18, 25), one month [25, 32), one to two months [32, 65), two to three months [65, 100), three to six months [100, 190), six to twelve months [190, 370), twelve to eighteen months [370, 550), eighteen to 24 months [550, 735), 24 to 36 months [735, 1080). Since early reported failure times were considered to have been fairly precisely recorded, early intervals were defined to be correspondingly short. After three months, the intervals were constructed in accordance with follow-up times. These intervals were extended by a few days beyond the scheduled follow-up date to include follow-up contacts that were made after the assigned follow-up date.

Additional data were collected from each participant to be related to failure times. Variables included in the analyses below were age; baseline cigarette consumption; baseline confidence in ability to quit; proportion of friends who smoke; number of previous quit attempts; clinic site; number of reasons to quit; indicators for health, cost and family reasons to quit; sex; number of years participant has smoked; and smoking treatment.

### 2.5.3. Assumptions on the Failure Time Distribution

Analyses from a similar comparative study (Koehler and McGovern, 1990) indicated that a separate Weibull limited failure population model (Meeker, 1987) for each treatment is reasonable for these data. This model assumes that the failure times for the failing subpopulation follow a Weibull distribution. A parameter is included that measures the relative size of the failing subpopulation. Meeker (1987) notes that precise estimates for the parameters in this model may not be obtained if less than 80% of those who will eventually fail are actually observed; according to McGovern (personal communication; School of Public Health, University of Minnesota), approximately 96-97% of the true failures had been observed by the end of the study.

Explanatory variables were included in the analyses by modeling the log of the Weibull scale parameter as a linear function of the explanatory variables. Exploratory analyses were performed on the data to determine which explanatory variables appeared to influence failure times. The LIFEREG procedure in SAS (1985) was used with a Weibull assumption and a log-linear model on the Weibull scale parameter. The procedure was applied to the recorded (not interval censored) failure and right censor times with most of the individuals who were abstinent at the end of the study omitted from the data set. The deletion of these participants was used to create a data set that represented the failing subpopulation only. The LIFEREG procedure assumes independence, but this was not of great concern since violations of the independence assumption generally affect estimated variances of the parameter estimators much more than the realized values of the parameter estimators.

Although the data set consisted of 915 individuals who actually made a quit attempt, complete data for all variables were available for only 88% of the individuals. An initial regression for each treatment was performed to select a smaller set of important variables. The variables selected from this initial pass were the proportion of friends who smoked, the baseline cigarette consumption, whether the participant had a cost reason to quit, and the baseline quit confidence. Further analyses indicated that baseline quit confidence, scaled from 0 (no confidence) to 100 (complete confidence), appeared to be the only important factor for the treatment failure time distributions. Thus, the log-linear model on the scale parameter for the analyses below was taken to be a function of baseline quit confidence for each treatment. There were 871 complete records for this model. Attributes of this data set, including the total number of participants and groups and the number of failures in each interval, are listed for each treatment in Table 2.1.

### 2.5.4. The Mean Model

The mean of a binary response for a particular treatment can be derived from a mixture of the distributions for the failing and abstaining subpopulations. Consider first the Weibull failure time distribution for the failing subpopulation. Using a log-linear relationship between the scale parameter and the explanatory variables, the survival function for failure times in this subpopulation is

$$S_f(t) = \exp\{ - [\exp(-\underline{X}'\underline{\beta})\ t]^{\eta} \} ,$$

where $\eta$ is the Weibull shape parameter and $\underline{\beta}$ is the vector of parameters associated with the explanatory variables $\underline{X}$. In the specific model used below, $\underline{X}' = (1, X_1)$, where $X_1$ contains the value for the baseline quit confidence, and

Table 2.1. Some attributes of the smoking cessation data set used to estimate the mean model parameters

| Attribute | Treatment | | |
|---|---|---|---|
| | ACS | ALA | LANDO |
| Number of Participants | 260 | 312 | 299 |
| Number of Groups | 33 | 31 | 33 |
| Average Quit Confidence | 64.1 | 63.6 | 66.7 |
| Number of Failures/ Censors in Interval | | | |
| [ 0, 4) | 48 / 0 | 48 / 0 | 38 / 0 |
| [ 4, 11) | 43 / 1 | 55 / 0 | 34 / 0 |
| [ 11, 18) | 28 / 0 | 21 / 0 | 18 / 0 |
| [ 18, 25) | 28 / 0 | 24 / 0 | 18 / 0 |
| [ 25, 32) | 16 / 0 | 23 / 0 | 28 / 0 |
| [ 32, 65) | 28 / 0 | 35 / 0 | 33 / 0 |
| [ 65, 100) | 11 / 0 | 14 / 0 | 20 / 0 |
| [100, 190) | 16 / 0 | 16 / 0 | 23 / 1 |
| [190, 370) | 5 / 0 | 12 / 0 | 12 / 1 |
| [370, 550) | 2 / 0 | 2 / 0 | 4 / 1 |
| [550, 735) | 0 / 0 | 5 / 0 | 0 / 1 |
| [735,1080) | 2 / 32 | 2 / 55 | 2 / 65 |
| Total | 227 / 33 | 257 / 55 | 230 / 69 |

$\underset{\sim}{\beta} = (\beta_0, \beta_1)'$. For the abstaining subpopulation, an individual never fails. Thus $S_a(t) \equiv 1$.

The hazard probability for the entire population during interval h can be calculated from the hazard definition,

$$\Pr\{\text{ failing in interval } h \mid \text{success up to } t_{h-1} \}$$
$$= \Pr\{ T \in [t_{h-1}, t_h) \} / \Pr\{ T > t_{h-1} \} .$$

Let $\phi$ be the proportion of failers in the population; i.e., the probability that an individual belongs to the failing subpopulation. The expression for the numerator can be derived as follows.

$$\Pr \{ T \in [t_{h-1}, t_h) \}$$
$$= \Pr \{ T \in [t_{h-1}, t_h) \mid \text{abstainer} \} \Pr\{ \text{abstainer} \}$$
$$+ \Pr \{ T \in [t_{h-1}, t_h) \mid \text{failer} \} \Pr\{ \text{failer} \}$$
$$= [S_a(t_{h-1}) - S_a(t_h)] (1 - \phi)$$
$$+ [S_f(t_{h-1}) - S_f(t_h)] \phi$$
$$= \phi [\exp\{ - [\exp(-\underset{\sim}{X}'\underset{\sim}{\beta}) t_{h-1}]^{\eta} \}$$
$$- \exp\{ - [\exp(-\underset{\sim}{X}'\underset{\sim}{\beta}) t_h]^{\eta} \}] .$$

Using the same mixture argument,

$$\Pr \{ T > t_{h-1} \}$$
$$= S_a(t_h) (1 - \phi) + S_f(t_h) \phi$$

$$= 1 - \phi\,[1 - \exp\{\, - [\exp(-\underline{X}'\underline{\beta})\; t_{h-1}]^{\eta}\,\}] \;.$$

Thus, the mean model for the response of individual i in group j during interval h is

$$\pi_{hij} = [\phi^{-1} - 1 + \exp\{\, - [\exp(-\underline{X}'_{hij}\underline{\beta})\; t_{h-1}]^{\eta}\,\}]^{-1}$$

$$\times\, [\exp\{\, - [\exp(-\underline{X}'_{hij}\underline{\beta})\; t_{h-1}]^{\eta}\,\}$$

$$- \exp\{\, - [\exp(-\underline{X}'_{hij}\underline{\beta})\; t_{h}]^{\eta}\,\}] \;.$$

The mean model parameters to be estimated from the data are $\phi$, $\eta$, $\beta_0$ and $\beta_1$.

### 2.5.5. The Correlation Model

Observations during the course of the study indicated that related individuals, for example married couples, tended to have more highly correlated failure times than unrelated individuals. Although cohesiveness sometimes increases in support groups over time, the short duration of the treatment program relative to the length of the study did not suggest that the correlations should be allowed to vary across time as well. Two classes of correlation were assumed to exist among the binary responses of group members to account for differences in relationships among participants. The correlation classes correspond to related pairs (married couples and pairs with the same last name) and unrelated pairs (all other pairs of individuals). This is a simplification of the true correlation structure; more levels of correlation could have been constructed if more information had been illicited from the subjects, such as which subjects in the group were friends or sharing living accommodations. The effect of this type of correlation structure is that groups with a higher proportion of related individuals tend to have higher "average" correlations among binary responses. The auxiliary covariance parameters to be estimated under this model are $\underline{\alpha} = (\rho_1, \rho_2)'$, where $\rho_1$ is the correlation coefficient for related pairs and $\rho_2$ is the correlation coefficient for unrelated pairs.

### 2.5.6. Consistent Initial Estimators for $\underline{\gamma}$ and $\underline{\alpha}$

Consistent initial estimates of $\underline{\gamma}$ were obtained using the maximum likelihood approach assuming independence among individuals. A FORTRAN program to estimate Weibull limited failure population models (Meeker, 1983) was modified to allow explanatory variables to be included in a log-linear model for the scale parameter. The likelihood accounted for interval and right censoring. Let

$$\delta_{1hij} = \begin{cases} 1 & \text{if individual } j \text{ in group } i \text{ failed during interval } h \\ 0 & \text{otherwise} \end{cases}$$

and

$\delta_{2hij} = 1$ if the failure time for individual j in group i is right censored during interval h

$= 0$ otherwise .

The likelihood used to obtain $\hat{\gamma}^{(0)}$ was

$$\prod_{h=1}^{k+1} \prod_{i=1}^{m_h} \prod_{j=1}^{n_{hi}} \phi\,[S_f(t_{h-1}) - S_f(t_h)]^{\delta_{1hij}}$$
$$\times\,[\phi\, S_f(t_{h-1}) + (1-\phi)]^{\delta_{2hij}} .$$

Estimates of $\gamma$ from this procedure for each treatment are listed in Table 2.2.

To estimate the two correlation coefficients for the covariance matrix, intra-group pairs of individuals were separated into two disjoint classes, related pairs and unrelated pairs. Pairs of participants were considered to be related if they were known to be married or if they shared a common last name. Estimator (5.7) in Chapter 5 was used to estimate the correlation coefficient corresponding to each correlation class. This estimator was selected based on the simulation results cited in Chapter 5.

Estimates for the pairwise correlation coefficients for each correlation class in each treatment are listed in Table 2.3. The high degree of correlation among related pairs relative to unrelated pairs for all treatments is evident from the estimates. The related correlation estimate for the ACS treatment is much higher than the estimates for the

Table 2.2. Initial estimates (and standard errors) of the mean model parameters for each smoking cessation treatment assuming independence among all observations

| Parameter | Treatment | | |
| --- | --- | --- | --- |
| | ACS | ALA | LANDO |
| Proportion of Failers | 0.873 (.020) | 0.816 (.022) | 0.776 (.024) |
| Shape | 0.822 (.044) | 0.738 (.037) | 0.831 (.043) |
| Intercept | 4.28 (.22) | 3.84 (.23) | 4.02 (.19) |
| Baseline Quit Confidence | -0.0069 (.0031) | 0.0038 (.0032) | 0.0033 (.0027) |

Table 2.3. Initial estimates of the correlation coefficients for related and unrelated pairs for each smoking cessation treatment

| Correlation Class | Treatment | | |
| --- | --- | --- | --- |
| | ACS | ALA | LANDO |
| Related Pairs | .687 | .249 | .133 |
| Unrelated Pairs | .028 | .032 | .017 |

other treatments. One possible interpretation of this result is that the lower number of group meetings in the ACS program does not provide individuals with as many tools for self-discipline, vehicles for positive reinforcement, and coping strategies for withdrawal stress, and this causes the success of related couples to be more highly linked for the ACS treatment than for the other treatments.

### 2.5.7. Results of Analyses and Discussion

Using the Gauss-Newton algorithm described in Section 2.3.2, the four mean model parameters and their standard errors were estimated for each treatment. Results are presented in Table 2.4. The estimated baseline quit confidence parameters and standard errors indicate that baseline quit confidence is not significant for any treatment.

Contrasts comparing the remaining parameters across treatments (the ACS versus the average of the ALA and LANDO parameters, and the ALA versus the LANDO parameter) were tested using two-sided t-tests. Based on the asymptotic normality of the parameter estimates, approximate t-tests were constructed by assuming that parameter estimates for different treatments were independent and that variances of the parameter estimates were homogeneous across treatments. Since four parameters were estimated in each model, the

Table 2.4. Mean model parameter estimates (with estimated standard errors) for each smoking cessation treatment from nonlinear least squares estimation accounting for intra-group correlations

| Parameter | Treatment ACS | Treatment ALA | Treatment LANDO |
|---|---|---|---|
| Proportion of Failers | 0.854 (.022) | 0.801 (.024) | 0.766 (.025) |
| Shape | 0.517 (.035) | 0.524 (.035) | 0.593 (.038) |
| Intercept | 3.59 (.27) | 3.54 (.28) | 3.77 (.29) |
| Baseline Quit Confidence | -0.0034 (.0036) | 0.0029 (.0040) | 0.0026 (.0040) |

degrees of freedom for the standard errors were 29 for the ACS and LANDO treatments and 27 for the ALA treatment so that the pooled estimate of variance had 85 degrees of freedom. Values for the t-statistics are listed in Table 2.5.

Test results for the estimated proportion of failers indicate that the more intensive ALA and LANDO treatments have higher success rates than the ACS program. No difference was detected between the ALA and LANDO programs for the relative size of the recidivist population. These patterns were also observed in analyses conducted by Lando et al. (1990).

The results also suggest that the shape and intercept parameters are nearly constant across all treatments. The value of the shape parameter is less than one, indicating that the hazard function declines monotonically over time. The shape of the hazard reflects the fact that the risk of failure is much higher in the early phases of smoking cessation.

These tests indicate that a limited failure population Weibull model with common shape and scale parameters across treatments and separate failing proportion parameters for each treatment may be appropriate for these data.

The estimated values for some parameters in Table 2.4 differ from the initial consistent estimates cited in

Table 2.5. Approximate t-statistics for contrasts of the mean model parameters comparing smoking cessation treatments

| | Contrast | |
|---|---|---|
| Parameter | ACS vs. Others | ALA vs. LANDO |
| Proportion of Failers | 2.42[a] | 1.04 |
| Shape | -0.94 | -1.33 |
| Intercept | -0.19 | -0.58 |

[a]Reject null hypothesis at $\alpha = .05$ on 85 degrees of freedom if $|t| > 1.99$.

Table 2.2. Estimates of the intercept and particularly the shape parameters are substantially lower when within-group correlations are considered. The effects of these changes in the parameter values on the failure time distribution offset one another to some extent, but the least squares estimates of the hazard functions for the three treatments are more skewed than the corresponding independence-based estimates. The estimated proportion of recidivists is also slightly smaller for each treatment when within-group correlations are included in the model; these changes are probably a result of the increased skewness in least squares estimate of the recidivist distribution.

The new methodology is expected to produce more accurate (i.e., usually larger) estimates of the standard errors for the parameter estimates by accounting for correlation among observations. Estimated standard errors for the least squares procedure were about 10% higher than those for the independence-based procedure for the estimated proportion of failers, and about 30% higher for the regression coefficients (Tables 2.2 and 2.4). However, estimated standard errors from the Gauss-Newton algorithm were about 10% lower for the shape parameter estimates than from the independence-based analysis. One possible explanation is that since the shape parameter is bounded below by zero, the variance estimate is related to the parameter estimate; hence the lower estimated standard errors from the Gauss-Newton algorithm may be a consequence of the lower shape estimate from this estimation method. The relatively small differences in estimated standard errors between the two estimation procedures are the result of very low estimated correlations for most pairs of responses in the data. In general, larger intra-group correlations will lead to larger differences in estimated standard errors between the independence-based and least squares procedures.

# 3. ANALYSES FOR EXACT FAILURE TIME DATA COLLECTED FROM INDEPENDENT GROUPS OF CORRELATED INDIVIDUALS

## 3.1 Introduction

In Chapter 2, a method is developed for analyzing commonly interval and possibly right censored failure time data from independent groups of correlated individuals. Another common form for failure time data arising from grouped individuals is exact time and right censored data. Methods currently available for analyzing this type of data are discussed Chapter 1. These approaches are limited in their application by restrictions placed on group size, intra-group correlations and/or the ability to incorporate explanatory variables into the analyses.

One alternative approach for analysis of exact failure time data for grouped individuals is to extend the nonlinear least squares estimation techniques described in Chapter 2 by expressing the observed failure time data as conditional binary variates. The binary variables are constructed from the observed failure and censor times using appropriately defined intervals for the exact time data. This approach permits large, variable group sizes and heterogeneous intra-group correlations. In addition, a wide variety of failure time distributions can be estimated, including distributions that depend on explanatory variables.

Another possible approach to analyzing exact time data from groups of correlated individuals is to employ methods related to the generalized estimating equation approach of Liang and Zeger (1986) using the observed failure and censor times. This method is a multivariate extension of quasilikelihood methods for generalized linear models and is closely related to multivariate nonlinear least squares estimation. Like the analyses based on conditional binary variates, this methodology allows for large and variable group sizes with a variety of correlations structures, and permits analysis of explanatory variable effects. However, right censored times are not easily incorporated into the estimating equations.

In this chapter, estimation methods using exact failure times and conditional binary variates are described and compared. An illustration of the conditional binary variate approach is presented using data from a toxicological study.

## 3.2 Analyses Based on Observed Exact Failure and Censor Times

Liang and Zeger (1986) and Zeger and Liang (1986) describe a multivariate extension of quasilikelihood methods for generalized linear models. Univariate quasilikelihood methods, described in McCullagh and Nelder (1983), are used to model the means of random variables as a known function of a linear combination of explanatory variables when the variance of the random variable is related to the mean. An underlying family of distributions that has a form related to the exponential family is assumed for the data. Explanatory variable effects are incorporated by taking the mean of this family to be a known function of a linear combination of the explanatory variables, $\underline{X}'\underline{\beta}$, where $\underline{X}$ is the vector of explanatory variables and $\underline{\beta}$ is the parameter vector to be estimated. Estimates of the mean model parameters can be obtained by solving the system of equations derived from setting the partial derivatives of the log-likelihood with respect to the elements of $\underline{\beta}$ equal to zero. For a set of independent observations, this system is defined by

$$\sum_i \underline{d}_i \, (T_i - \mu_i) \, / \, v(\mu_i) = \underline{0} \ ,$$

where $T_i$ is the realized value for observation i, $\mu_i$ is the mean of $T_i$ depending on $\underline{\beta}$, $v(\mu_i)$ is the variance of $T_i$ and a known function of $\mu_i$, and $\underline{d}_i = \partial\mu_i/\partial\underline{\beta}$.

Liang and Zeger (1986) extend the univariate quasilikelihood approach to the case where groups of observations are correlated. They make the same assumption on the marginal distribution of the responses as in univariate quasilikelihood, but the full multivariate likelihood is not explicitly defined. Instead, a correlation matrix is constructed describing the relationships among the correlated observations. The system of equations for obtaining estimates of $\underline{\beta}$ for independent groups of correlated observations is defined by

$$\sum_i D_i' \, V(\underline{\mu}_i)^{-1} \, (\underline{T}_i - \underline{\mu}_i) = \underline{0} \, ,$$

where $\underline{T}_i$ is the vector of realized values for group i, $\underline{\mu}_i$ is the mean of $\underline{T}_i$ depending on $\underline{\beta}$, $V(\underline{\mu}_i)$ is the covariance matrix for $\underline{T}_i$ and a known function of $\underline{\mu}_i$, $\underline{T}_i$ is the vector of correlated observations, and $D_i = \partial\underline{\mu}_i/\partial\underline{\beta}'$. If the variance of the j-th element of $\underline{T}_i$ is defined to be $g(\mu_{ij})$, $V_i$ can be decomposed such that

$$V_i(\underline{\mu}_i) = A_i^{1/2} \, R_i(\underline{\alpha}) \, A_i^{1/2} \, ,$$

where $A_i = \mathrm{diag}\{\, g(\mu_{ij}) \,\}$ and $R_i(\underline{\alpha})$ is the working correlation matrix describing the relationships among observations in group i. $R_i(\underline{\alpha})$ is said to be a "working" correlation matrix because it does not need to be correctly specified for the asymptotic properties of the estimators described below to hold. Hence if nothing is known about the intra-group corrections, $R_i(\underline{\alpha})$ may be taken to be the identity matrix.

Liang and Zeger show that these equations produce consistent and asymptotically normal parameter estimates regardless of the specification of the correlation matrix $R_i(\underline{\alpha})$. They also provide a consistent estimator of the covariance matrix of the parameters based on the observed covariance matrix of $\underline{T}_i$. However, these results require $\mu_{ij}$ to be a function of a linear combination of the parameters. For many failure time models, the mean may be intrinsically nonlinear in the parameters. When this is the case, multivariate nonlinear least squares estimation can be used to obtain asymptotically normal parameter estimates by inserting the observed failure times and the appropriate mean models and covariance matrices into the equations for the Gauss-Newton algorithm. However, a sufficient condition for the asymptotic properties of Gauss-Newton estimators is the correct specification of the covariance matrix of $\underline{T}_i$ (see Chapter 4).

Analyzing uncensored exact failure time data via generalized estimating equations or nonlinear least squares estimation is relatively straightforward. Suppose that uncensored exact failure times, $T_{ij}$, are available from independent groups of correlated individuals, where $i = 1, 2, \ldots, m$ groups and $j = 1, 2, \ldots, n_i$ individuals in group $i$. Several types of assumptions may be used to model the mean of $T_{ij}$. For example, any parametric distribution may be used to derive a mean model for $T_{ij}$. If an estimate of the hazard function is of interest, a parametric model is easier to work with. However, a nonparametric model can be used to derive the mean as a function of the steps defining the nonparametric hazard function. To incorporate explanatory variables, location-scale distributions (Lawless, 1980) can be used to model the log failure times with a linear function of the covariates. The variance of $\log(T_{ij})$ is modeled in accordance with the underlying distribution. If effects of the explanatory variables in a proportional hazards setting are of primary interest, an explicit model for the baseline hazard is required. Hence, Cox's semi-parametric proportional hazards cannot be used since the mean model for $T_{ij}$ depends on an unspecified hazard function.

Consider the case of exact failure time data that include right censored times. Such data are comprised of two types of random variables corresponding to failure times and censor times. To implement either the generalized estimation equation or nonlinear least squares estimation approach, means and variances are required for the censor times as well as for the uncensored failure times. It may be difficult to choose a model for the means and variances of randomly right censored times due to lack of information; i.e., since the failure times represent censored censor times, most of the censor times are themselves censored.

In many instances, censored observations may not contribute to estimation of the failure time distribution parameters. For example, in studies that are terminated at a predetermined time, the mean censor time for truncated observations is the time of truncation and the variance and covariances of this time are zero. Since the corresponding elements of $V(\underline{\mu}_i)$ and $(\underline{T}_i - \underline{\mu}_i)$ are zero, the deletion of these observations is necessary. Hence truncated observations do not contribute to estimation of the failure time distribution. Regardless of the type of censoring, under the usual assumption of independence between the failure and censoring mechanisms, the mean models for the failure and censoring times cannot be functionally related. Hence the parameters for the censoring mean model are nuisance parameters, and unlike most other failure time analyses, censored data will not contribute to the estimation of the failure time distributions.

An alternative method of approaching right censored data using the generalized estimation equation approach is to return to the marginal distribution assumption and determine the appropriate univariate contributions to the likelihood, $S(T_{ij})$, where S is the survival function for $T_{ij}$. However, the log-likelihood contribution from $\log[S(t)]$ is unlikely to have a form that would allow censored observations to be conveniently incorporated into the generalized estimation equations.

### 3.3 Analyses Based on Conditional Binary Variables

Although observed failure times are referred to as exact, it is rare that "exact" times are actually observed. More often failure times are recorded to the nearest unit, such as to the nearest day or hour. A censoring interval of unit length is implicitly defined in this process, which is typically a small length relative to the duration of the study. Consequently the data may be viewed as censored according to a common interval censoring scheme, and the conditional binary variable approach described in Chapter 2 can be used to estimate parameters in the failure time distribution.

Without loss of generality, define the intervals to be of unit length (data can be rescaled if necessary). The disjoint set of intervals covering the study is $\{[t_h, t_{h-1}):$ $h = 1, 2, \ldots, k\} = \{[h-1, h): h = 1, 2, \ldots, k\}$, where k is the largest observed failure or right censored time. Recall from Section 2.2 in Chapter 2 that the observed failure times for individuals in group i can be expressed as a set of conditional binary response vectors for each interval $\{\underline{Y}_{hi} : h = 1, 2, \ldots, k \text{ intervals}\}$. The elements of this vector, $Y_{hij}$, are defined from the observed failure or censor time $T_{ij}$ for individual j in group i as follows:

$$\begin{aligned} Y_{hij} &= 1 \quad \text{if observed failure time } T_{ij} \in [h-1, h) \\ &\qquad\qquad \text{given } T_{ij} \geq h-1 , \\ &= 0 \quad \text{if observed failure or censor time } T_{ij} \geq h \\ &\qquad\qquad \text{given } T_{ij} \geq h-1 , \end{aligned}$$

where $i = 1, 2, \ldots, m_h$ groups in interval h and $j = 1, 2, \ldots, n_{hi}$ individuals in the risk set (individuals surviving up to h-1) minus the censor set (individuals censored in [h-1, h) ) of group i during interval h. $Y_{hij}$ remains undefined if the failure occurred in a previous time interval or the individual was censored prior to time h.

Given $S_{ij}(t)$, the survival function for individual j in group i, the mean of $Y_{hij}$ for the unit-interval censored data is a hazard probability calculated from

$$\pi_{hij} = 1 - [S_{ij}(h) \,/\, S_{ij}(h-1)] \; . \tag{3.1}$$

The mean of $\underline{Y}_{hi}$ is defined to be $\underline{\pi}_{hi}$, whose elements are the hazard probabilities $\pi_{hij}$ corresponding to individual j in group i during interval h. The covariance matrix of $\underline{Y}_{hi}$, denoted $V_{hi}$, is a function of $\underline{\pi}_{hi}$ as well as a function of auxiliary parameters such as correlation coefficients describing the correlation structure within group i during interval h (see Section 2.2 in Chapter 2).

Multivariate nonlinear least squares estimation based on a Gauss-Newton algorithm can be used to estimate the mean model (i.e., failure time distribution) parameters. When the Gauss-Newton iterations are initiated with consistent estimates of the mean model and covariance matrix parameters, the estimated parameters can be shown to be jointly asymptotically normal under mild regularity conditions (Chapter 4).

In the event that exact failure times are actually observed, an approximate mean can be derived. Since the hazard function for individual j from group i, $\lambda_{ij}(t)$, is defined to be

$$\lambda_{ij}(t) = \lim_{\Delta \to 0} \Delta^{-1} \Pr\{ T_{ij} \in [t, t+\Delta) \mid T_{ij} \geq t \} ,$$

for small $\Delta$

$$\Pr\{ T_{ij} \in [t, t+\Delta) \mid T_{ij} \geq t \} \cong \Delta \, \lambda_{ij}(t) .$$

If $\Delta = 1$, the smallest unit of time encountered in the study, is small relative to the lifetime of the individual, the mean for the conditional binary variate $Y_{hij}$ is approximately

$$\pi_{hij} \cong \lambda_{ij}(h) \ . \tag{3.2}$$

Results from mean models (3.1) and (3.2) should be similar since

$$1 - [S_{ij}(h) \ / \ S_{ij}(h-1)]$$
$$= \Pr\{ \ T_{ij} \in [t, \ t+\Delta) \ | \ T_{ij} \geq t \ \}$$
$$\cong \Delta \ \lambda_{ij}(t) \ .$$

The conditional binary variable approach can be used with many types of failure time models. As in the approach based on the observed exact times, exceptions include a proportional hazards model with an unspecified baseline hazard function. The form of the assumed correlation matrix is limited only by the ability to obtain a consistent initial estimate of the covariance matrix parameters (see Section 2.3.4 in Chapter 2 for a discussion of acceptable models). In addition, unlike the approach based on exact failure times described in Section 3.2, censored data always contribute to the estimation of the failure time distribution parameters when binary variables are used. Hence, nonlinear regression based on conditional binary variables uses more of the information contained in the data than the analyses described in Section 3.2 based on exact failure and censor times.

## 3.4. Application of the Conditional Binary Variable Approach to Collaborative Behavioral Toxicological Study Data

### 3.4.1. Study Description

The Collaborative Behavioral Toxicological Study was designed by the National Center for Toxicological Research to study the reliability and sensitivity of behavioral testing methods to the effects of prenatal chemical exposure in rats. A complete description of the study objectives, design and analyses are available in a series of articles in Neurobehavioral Toxicology and Teratology, 1985, Volume 7. In this section, the focus is on the effects of methylmercuric chloride on the time to occurrence of two developmental landmarks, eye opening and first incisor eruption. A brief summary of the design for this part of the study follows.

The study was conducted at six different laboratories in the US. At each lab, four replicates of the methylmercuric chloride experiment were run. The design for each replicate included four treatment levels, with four rat litters assigned to each treatment. Litters were culled to contain two males and two females. Occasionally litter sizes were reduced to less than four pups due to death or unavailability of a pup of a particular sex. In eight cases, a whole litter was lost from a replication. For both the eye opening and incisor eruption data, there were 376 litters and 1476 pups.

To avoid excessive censoring due to death, the high dose level for methylmercuric chloride was selected to mitigate problems with infant mortality. In particular, for the developmental landmarks, the high dose was selected so that the average day of occurrence was not shifted by more than one day from the control pups. Treatments included untreated control, vehicle control (nitrogen-purged distilled water), 2.0 mg methylmercuric chloride / kg body weight, and 6.0 mg methylmercuric chloride / kg body weight. Doses were administered to pregnant females on gestation days six through nine.

Rat pups were inspected daily after birth for the occurrence of eye opening and incisor teeth eruption. Although this type of data is generally considered to be exact, this inspection schedule is more accurately described as interval censored, with intervals of length one day. The occurrence of both landmarks was noted for all rats retained in the study; hence no right censoring is present in the data.

The objective of the following analyses is to estimate the effects of treatments, labs and sex on the event time distributions for eye opening and incisor eruption.

### 3.4.2. Exploratory Analyses

Event time data for each landmark variable were examined to determine an adequate event time model. The number of events occurring for each variable on each day following birth are listed in Table 3.1. These data suggest that the eye opening event time distribution is fairly symmetric, but that the data for incisor eruption are skewed. It is also clear that a shift parameter is necessary to set the lower bound of each event time distribution at an appropriate positive value. The shift parameter for each variable was selected to be the minimum day of occurrence minus one (11 days for eye opening, 7 days for incisor eruption). Probability plots for shifted Weibull, lognormal and log-logistic distributions (without explanatory variable effects) suggest that a Weibull model most closely fits the distribution of both event time variables.

PROC LIFEREG (SAS, 1985) was used to estimate a Weibull distribution with the scale parameter modeled as a log-linear function of four treatment indicator variables, five lab indicator variables, and a sex indicator variable. For both event time variables, treatment and lab effects were significant in the preliminary analyses, but sex effects were not. Residual plots indicated that the shifted Weibull distribution with the scale parameter modeled as a

Table 3.1. Frequency of rat pups observed to exhibit developmental landmark on each day following birth for eye opening and incisor eruption

| Eye Opening | | Incisor Eruption | |
|---|---|---|---|
| Pup Age (days) | Frequency | Pup Age (days) | Frequency |
| 12 | 4 | 8 | 17 |
| 13 | 85 | 9 | 70 |
| 14 | 422 | 10 | 305 |
| 15 | 601 | 11 | 544 |
| 16 | 308 | 12 | 363 |
| 17 | 51 | 13 | 135 |
| 18 | 5 | 14 | 33 |
| | | 15 | 7 |
| | | 16 | 2 |

log-linear function of explanatory variables provides a fairly reasonable fit for both eye opening and incisor eruption (Figure 3.1).

### 3.4.3. Mean Model

Since the exploratory analyses indicated that a Weibull distribution with a log-linear model on the scale parameter is appropriate for these data, the mean model for the binary response for individual j in group i during interval h was defined to be

$$\pi_{hij} = 1 - \exp\{\ \exp(-\underline{X}'_{ij}\underline{\beta})^{\eta}\ [(h-\delta)^{\eta} - (h-1-\delta)^{\eta}]\ \}\ ,$$

where $\eta$ is the Weibull shape parameter, $\delta$ is the known shift parameter, $\underline{X}_{ij}$ is the vector of explanatory variables for individual j in group i, and $\underline{\beta}$ is the parameter vector for the explanatory variables. More specifically, the vector of explanatory variables was defined to be

$$\underline{X} = (X_1, X_2, \ldots, X_9, X_{10})'$$

= (indicator for untreated control, indicator for vehicle control, indicator for low methylmercuric chloride dose, indicator for high methylmercuric chloride dose, indicator for lab 2, ..., indicator for lab 6, indicator for males)' .

PROC LIFEREG in SAS (1985) was applied to the observed event times to obtain the consistent estimates of the

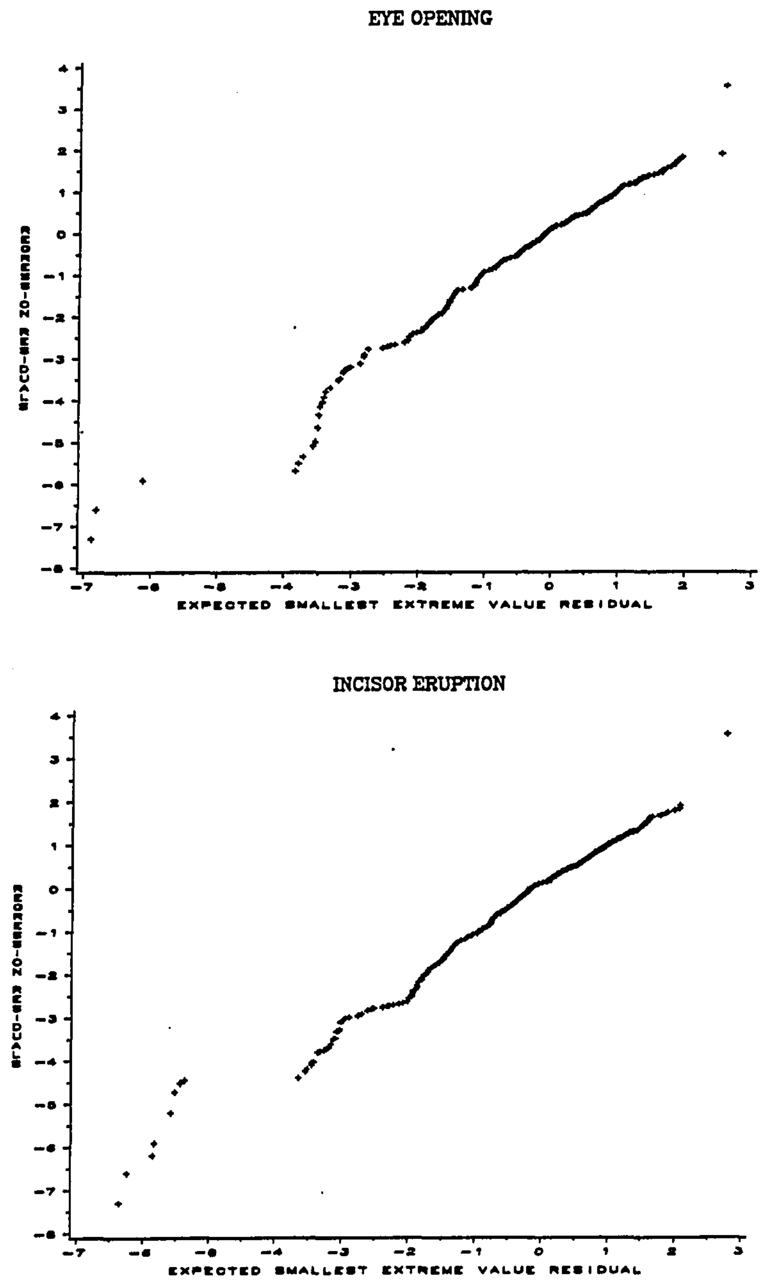


Figure 3.1. Smallest extreme value probability plot for the residuals from the shifted Weibull model with log-linear model on the scale parameter for day of eye opening and incisor eruption.

parameters in the above Weibull model that were used to initiate the Gauss-Newton algorithm. These initial estimates and estimated standard errors, calculated under the assumption of independence for all individuals, are listed in Table 3.2 for both eye opening and incisor eruption.

### 3.4.4. Covariance Matrix Model

The auxiliary covariance matrix parameters for the two event time variables consist of correlation coefficients describing the intra-group correlation structure. For eye opening and incisor eruption, it is possible that male-male (mm), male-female (mf), and female-female (ff) correlations may differ. Although estimator (5.7) in Chapter 5 is a consistent estimator for this model, it produced an estimated correlation exceeding one for the incisor eruption female-female correlation. Consequently, estimator (5.8) in Chapter 5 was used. Since the only factor that changes in the means for group members is gender, estimator (5.8) is consistent for male-male and female-female correlations. Furthermore, the parameter estimate for the sex indicator variable is quite small relative to the estimates of the other parameters, so estimator (5.8) is nearly consistent for the male-female correlation. The values of the estimated correlation coefficients using estimator (5.8) are

Table 3.2. Initial estimates (and standard errors) assuming independence among all individuals for eye and incisor eruption event time distribution parameters

| | Event Time Variables | |
|---|---|---|
| Parameter | Eye Opening | Incisor Eruption |
| Shape | 4.851<br>(.020) | 4.060<br>(.019) |
| Untreated Control | 1.454<br>(.017) | 1.672<br>(.020) |
| Vehicle Control | 1.477<br>(.017) | 1.705<br>(.020) |
| Low Dose | 1.410<br>(.017) | 1.649<br>(.020) |
| High Dose | 1.399<br>(.017) | 1.565<br>(.021) |
| Lab 2 | 0.029<br>(.018) | -0.145<br>(.022) |
| Lab 3 | 0.132<br>(.019) | -0.093<br>(.022) |
| Lab 4 | -0.137<br>(.019) | -0.129<br>(.022) |
| Lab 5 | 0.006<br>(.019) | -0.208<br>(.022) |
| Lab 6 | -0.074<br>(.019) | -0.243<br>(.022) |
| Male | 0.014<br>(.011) | 0.016<br>(.012) |

Table 3.3. Estimated correlations for each gender pair for eye opening and incisor eruption

| Correlation Class | Event Time Variable | |
|---|---|---|
| | Eye Opening | Incisor Eruption |
| Male-Male | .602 | .242 |
| Male-Female | .575 | .131 |
| Female-Female | .541 | .680 |

listed in Table 3.3.

Estimated correlation coefficients for each class of pairs are large and relatively homogeneous for eye opening. For incisor eruption, correlations for different sex pairs are quite different, with a particularly large female-female correlation relative to the estimates for the other correlation classes. The large intra-litter correlations, particularly for eye-opening, may indicate a strong genetic component in the occurrence of developmental landmarks.

### 3.4.5. Results and Discussion of Gauss-Newton Estimation

Parameters of the event time distributions for eye opening and incisor eruption were estimated using the Gauss-Newton algorithm described in Chapter 2. Results are presented in Table 3.4.

From the magnitude of the parameters, it is clear that treatments have the greatest effect on the day of eye opening and incisor eruption. Lab effects are not as strong, although most lab parameters are significantly different from zero. Sex of the pup has smaller effects on the event times; for incisor eruption, the parameter for the male indicator variable is not significantly different from zero.

Two-sided t-tests for contrasts of the treatment parameters were constructed to test specific hypotheses

Table 3.4. Nonlinear least squares estimates (and standard errors) accounting for intra-litter correlations for eye opening and incisor eruption event time distribution parameters

| | Event Time Variables | |
|---|---|---|
| Parameter | Eye Opening | Incisor Eruption |
| Shape | 3.975<br>(.101) | 2.973<br>(.057) |
| Untreated Control | 1.299<br>(.025) | 1.516<br>(.023) |
| Vehicle Control | 1.309<br>(.024) | 1.513<br>(.024) |
| Low Dose | 1.233<br>(.026) | 1.592<br>(.024) |
| High Dose | 1.243<br>(.026) | 1.403<br>(.026) |
| Lab 2 | 0.034<br>(.028) | -0.143<br>(.026) |
| Lab 3 | 0.150<br>(.024) | -0.074<br>(.024) |
| Lab 4 | -0.131<br>(.033) | 0.062<br>(.022) |
| Lab 5 | 0.017<br>(.028) | -0.196<br>(.027) |
| Lab 6 | -0.063<br>(.030) | -0.246<br>(.028) |
| Male | 0.0138<br>(.0068) | -0.024<br>(.014) |

regarding treatment effects. Tests comparing the two types of controls, the two methylmercuric chloride doses, and the average of the control and methylmercuric chloride dose parameters were calculated and are listed in Table 3.5. Since there were 365 degrees of freedom (376 groups minus 11 estimated parameters), calculated t-values were compared to a normal table.

For both landmark variables, no difference exists between the untreated and treated controls. Similar results were obtained by Buelke-Sam et al. (1985) using an analysis of variance model for event times that accounted for the nested structure of the experiment.

For eye opening, there is no difference between effects of the low and high methylmercuric chloride doses, but the average of the two dose parameters is significantly smaller than the average of the control parameters. These results imply that either dose of methylmercuric chloride leads to a smaller scale parameter for the eye opening event time distribution, effectively shifting the eye opening distribution towards earlier ages. Buelke-Sam et al. (1985) also found that the average day of eye opening was earlier for methylmercuric chloride treated pups.

There is no difference between the average of the control parameters and the average of the dose parameters for the incisor eruption distribution. However, a

Table 3.5. Calculated t-values for contrasts of eye opening and incisor eruption treatment parameters

| | Event Time Variable | |
|---|---|---|
| Contrast | Eye Opening | Incisor Eruption |
| Untreated versus Vehicle Control | -0.56[a] | 0.11 |
| Low versus High Methylmercuric Chloride Dose | -0.43 | 7.72 |
| Average Control versus Average Dose | 4.24 | 1.07 |

[a]Reject null hypothesis of no difference at $\alpha = .05$ with 365 degrees of freedom if $|t| > 1.96$.

significant difference is evident between the low and high dose parameters. The low dose parameter is larger than the high dose parameter (and the controls). This indicates that low methylmercuric chloride doses depress and high doses accelerate incisor development. This contradicts previous analyses by Buelke-Sam et al. (1985), who found that the effect of both doses of methylmercuric chloride is to accelerate the process of incisor eruption.

In terms of the objectives of the study, it is clear that treatment effects are stronger than the other effects in the experiment. This is a desirable result indicating the relatively low sensitivity of the landmark variables to laboratory differences. Also, for both event time variables, gender does not appear to play a very important role in treatment effects, which may enable experimenters to use rats of only one sex in determining methylmercuric chloride effects on developmental landmarks. If it is legitimate to restrict the experiment to one sex and intra-sex correlations differ for males and females, lower standard errors for the estimated parameters of the event time distribution may be achieved by choosing pups belonging to the gender associated with lower intra-sex correlations.

Clear interpretation of the methylmercuric chloride dose effects is enhanced by that fact that the untreated and vehicle controls do not differ in their effects. Hence dose effects can be attributed to administrations of methylmercuric chloride rather than possible side effects of the vehicle used to inject the methylmercuric chloride into pregnant females.

Comparing results of the independence Weibull regression with results of the binary variable regression accounting for intra-litter correlations indicates that as expected, in nearly all cases, estimated standard errors for the binary variable regression are larger. The increase is largest for the shape parameter (500% for eye opening, 300% for incisor eruption). Increases in estimated standard errors for treatment and lab parameters are smaller for both eye opening (50%) and incisor eruption (10-30%). These results imply that failure to account for intra-group correlations leads to overstatement of the significance of the treatment comparisons. For the eye opening male indicator parameter, the estimated standard error is cut in half; for incisor eruption, this standard error remains about the same. Since gender is a within-group effect, it is possible that accounting for intra-group correlations reduces the estimated standard error for the sex parameter. As expected, the higher intra-group correlations observed in the eye opening data lead to larger increases in the standard errors for eye opening parameters relative to the incisor eruption standard errors.

For both event time variables, the lab and sex parameter estimates appear to be quite similar for both the independence-based and least squares estimates. However, shape and treatment parameter estimates shift significantly for both eye opening and incisor eruption. In the case of incisor eruption, these changes led to a different ordering of the treatment effects. The shift in the shape and treatment parameters and the large standard errors for the shape parameter may be partially due to the representation of the data as interval censored for the Gauss-Newton estimation rather than as exact times for the independence-based estimation.

# 4. ASYMPTOTIC PROPERTIES OF MULTIVARIATE NONLINEAR LEAST SQUARES ESTIMATORS FOR FAILURE TIME DATA COLLECTED FROM INDEPENDENT GROUPS OF CORRELATED INDIVIDUALS

## 4.1. Introduction

Correlation among individuals is frequently encountered in survival data. One common form of correlation arises when individuals are arranged in groups, for example as pups in a litter, or as patients treated by a particular physician or medical center. Methodologies developed to analyze survival data with this type of structure have generally been limited to small group sizes, often involve restrictive assumptions on the correlation structure within and across groups, and/or are unable to incorporate explanatory variable information (see Chapter 1).

Chapters 2 and 3 describe a new method that is appropriate for interval censored or exact failure time data collected from independent groups of correlated individuals. The method accommodates large groups sizes with heterogeneous correlation structures, and allows for the use of explanatory variables. The analyses rely on representing the failure times as conditional binary variables indicating the failure of an individual during a specified interval given success in the previous interval. For each time interval, a separate vector of binary responses is constructed for each group, which has a mean whose elements are hazard probabilities and are thus functions of the failure time distribution parameters. The covariance matrix for the binary response vector is a function of the mean vector and parameters describing the correlation structure among the elements of the observed response vector. The parameters of the failure time distribution are estimated by assuming that the response vector is equal to the mean vector plus an error vector, and using multivariate nonlinear least squares techniques.

Nonlinear least squares estimators based on the Gauss-Newton algorithm are considered in this chapter. Early work in this area was performed by Jennrich (1969), who derived the asymptotic properties of nonlinear least squares estimators for univariate models. Many extensions of this work are outlined in Gallant (1987), including methods for multivariate nonlinear models.

For multivariate problems, nonlinear least squares estimators are typically based on response vectors of constant dimension, and a weighted residual sum of squares is minimized to obtain parameter estimates for the nonlinear model. In contrast, for the conditional binary response vector formulation, the residual sum of squares is summed over intervals and groups, and the length of the response vectors that make up the weighted squared deviates varies over intervals and groups.

The purpose of this chapter is to define a multivariate nonlinear least squares estimator that is appropriate for grouped survival data, and to outline sufficient conditions under which this estimator is asymptotically normal. After introducing a general model for the survival data, an estimator for the case of known covariance matrices is defined. For the more common case of unknown covariance matrices, estimators based on the Gauss-Newton algorithm are developed and shown to be asymptotically normal under mild regularity conditions as the number of groups becomes large.

## 4.2. Data Formulation and Model

Suppose that the data consist of observations from m independent groups of individuals and that failure or censor times for each individual are observed to fall into one of $k+1 < \infty$ disjoint intervals, $\{[t_{h-1}, t_h) : h = 1, 2, \ldots, k+1; t_0 = 0; t_{k+1} = \infty\}$, where time $t_k$ represents the end of the observation period for the study. Let $n_{hi}$ be the number of individuals in the risk set for group i at time $t_{h-1}$ minus the number of individuals who are censored during interval h, and let $m_h$ denote the number of groups in interval h with $n_{hi} > 0$. Note that $n_{k+1,i} = 0$ for all i and $m_{k+1} = 0$. Define $Y_{hij}$ such that

$$\begin{aligned} Y_{hij} &= 1 \quad \text{if individual } j \text{ in group } i \text{ fails during interval } h \text{ given success up to } t_{h-1}, \\ &= 0 \quad \text{if individual } j \text{ in group } i \text{ succeeds during interval } h \text{ given success up to } t_{h-1}, \end{aligned}$$

where $i = 1, 2, \ldots, m_h$ groups and $j = 1, 2, \ldots, n_{hi}$ individuals. $Y_{hij}$ is not defined for any individual who is censored during interval h or has failed or been censored before $t_{h-1}$. This implies that $Y_{k+1,ij}$ is undefined for all i and j, so that the sums over h in the following sections end at $h = k$.

Each conditional binary variable, $Y_{hij}$, follows a Bernoulli distribution with the mean given by hazard probability $\pi_{hij}$ and variance $\pi_{hij}(1 - \pi_{hij})$. The covariance of two observations in the same group and interval is

$$\rho_{hijj'} [\pi_{hij}(1 - \pi_{hij}) \, \pi_{hij'}(1 - \pi_{hij'})]^{1/2}$$

where $\rho_{hijj'}$ is the correlation between the two responses. Because groups are independent and responses in different intervals are conditioned on previous responses, observations from different groups and different intervals have zero covariance. Examples of models for $\pi_{hij}$ are described in Chapter 2.

An $n_{hi}$ x 1 observed response vector,

$$\underline{Y}_{hi} = \{Y_{hi1}, Y_{hi2}, \cdots Y_{hin_{hi}}\},$$

can be constructed for each group and each interval with mean $\underline{\pi}_{hi}$ and covariance matrix $V_{hi}$. The elements of

$$\underline{\pi}_{hi} = \{\pi_{hi1}, \pi_{hi2}, \cdots \pi_{hin_{hi}}\}$$

are functions of the parameters and possibly explanatory variables that define the underlying failure time distribution. The elements of $V_{hi}$ are functions of $\underline{\pi}_{hi}$ and additional parameters describing the correlation structure among observations for group i during interval h.

For the purposes of this chapter, $\underline{\pi}_{hi}$ and $V_{hi}$ will be written as explicit functions of the failure time distribution parameters, $\underline{\gamma}$, explanatory variables, $\underline{X}_{hi}$, and correlation parameters, $\underline{\alpha}$. $\underline{\pi}_{hi}$ and $V_{hi}$ will be denoted $\pi(\underline{\gamma},\underline{X}_{hi})$ and $V(\underline{\theta},\underline{X}_{hi})$, respectively, where $\underline{\theta} = (\underline{\gamma}', \underline{\alpha}')'$.

The model for $\underline{Y}_{hi}$ is assumed to be

$$\underline{Y}_{hi} = \underline{\pi}(\underline{\gamma},\underline{X}_{hi}) + \underline{e}_{hi}, \qquad (4.1)$$

where $\underline{\gamma}$ is an s x 1 vector of fixed, unknown parameters belonging to the parameter space $\Gamma$; $\Gamma$ is a compact subset of $\mathbb{R}^s$; $\underline{X}_{hi} = (\underline{X}'_{hi1}, \ldots, \underline{X}'_{hin_{hi}})'$ is an $rn_{hi}$ x 1 vector belonging to a compact subset of $\mathbb{R}^{rn_{hi}}$ containing the r x 1 explanatory variable vectors associated with each of the $n_{hi}$

individuals contributing to the hi-th observed response vector; $\underline{\pi}$ is an $n_{hi}$ x 1 vector whose elements are continuous functions from $\Gamma \times \mathbb{R}^{rn_{hi}}$ into [0,1] with continuous and uniformly bounded first and second and continuous third partial derivatives with respect to $\underline{\gamma}$; the $\underline{e}_{hi}$ are independent across h and i with mean $\underline{0}$ and nonsingular covariance matrix $V(\underline{\theta},\underline{X}_{hi})$ for $\underline{\theta} = (\underline{\gamma}', \underline{\alpha}')$; $\underline{\alpha}$ is a u x 1 vector of fixed, unknown parameters belonging to parameter space $\Phi$; and $\Phi$ is a compact subset of $\mathbb{R}^u$. Let $\Theta = \Gamma \times \Phi$ denote the parameter space for (s+u) x 1 fixed, unknown parameter vector, $\underline{\theta}$. Note that $\Theta$ is a compact subset of $\mathbb{R}^{s+u}$.

## 4.3. Preliminary Convergence Theorems

Two theorems regarding convergence of means of matrix products as the number of groups, m, tends to infinity are required to develop the asymptotic properties of the nonlinear least squares estimators for model (4.1). Several assumptions are required to prove these and other theorems in this chapter.

It is convenient to partition the response vectors for each value of h into groups whose vectors are of equal dimension, $\delta$. The first assumption states that there exists a finite maximum group size, n, for the population of groups. This provides a maximum length for the data vectors and implies that $\delta$ ranges from 1, 2, ..., n.

*Assumptions A4.1.* There exists a maximum group size, $n < \infty$, for the population of groups.

Let $m_{h\delta}$ denote the number of response vectors that have length $n_{hi} = \delta$ during interval h. Note that

$$m_h = \sum_\delta m_{h\delta} .$$

Both $m_h$ and $m_{h\delta}$ are assumed to be positive. For a given h, it is assumed that the proportion of response vectors with dimension $\delta$ converges to a constant as $m_h$ gets large.

*Assumption A4.2.* For each value of h and $\delta$,

$$\lim_{m_h \to \infty} m_h^{-1} m_{h\delta} = \Delta_{h\delta} ,$$

where $0 \le \Delta_{h\delta} \le 1$ and $\sum_\delta \Delta_{h\delta} = 1$ .

This assumption implies, for example, that for any interval h, the relative proportion of the sampled data vectors for a given length does not fluctuate unstably as the number of groups with $n_{hi} > 0$ increases. A condition is also needed indicating that a change in the limit index from $m_h$ to $m_{h\delta}$

can occur without altering the limit of matrices averaged over groups.

*Assumption A4.3.* Let $\{A_{hi}\}$ be a sequence of p x q matrices of real numbers. For each value of h and $\delta$,

$$\lim_{m_h \to \infty} m_{h\delta}^{-1} \sum_{i:n_{hi}=\delta} A_{hi} = \lim_{m_{h\delta} \to \infty} m_{h\delta}^{-1} \sum_{i:n_{hi}=\delta} A_{hi}$$

whenever the right hand side limit exists.

A similar set of conditions is required for the behavior of $m_h$ in relation to m, with stricter bounds on the limiting proportion of groups remaining in interval h.

*Assumption A4.4.* For each value of h,

$$\lim_{m \to \infty} m^{-1} m_h = \Delta_h \ ,$$

where $0 < \Delta_h < 1$ and $\sum_h \Delta_h = 1$ .

*Assumption A4.5.* Let $\{A_{hi}\}$ be a sequence of p x q matrices of real numbers. For each value of h,

$$\lim_{m \to \infty} m_h^{-1} \sum_{i=1}^{m_h} A_{hi} = \lim_{m_h \to \infty} m_h^{-1} \sum_{i=1}^{m_h} A_{hi}$$

whenever the right hand side limit exists.

The following assumption places a restriction on the vectors of explanatory variables.

*Assumption A4.6.* For each value of h and $\delta$, the empirical distribution function of an $r\delta$ x 1 vector of explanatory variables, denoted $F_{m_{h\delta}}$, converges to some distribution function $F_\delta$ as $m_{h\delta}$ tends to infinity.

This condition insures, for example, that the sample values of the explanatory variable vectors cannot oscillate indefinitely over certain areas of the explanatory vector space in a way that does not provide increasing information as the number of groups is increased. For survival data, assumption A4.6 is satisfied for fixed explanatory variable vectors that appear with some probability specific to each h and $\delta$, or when the $\{\underline{X}_{hi}\}$ are a random sample from $F_\delta$. This assumption also admits the use of time-dependent explanatory variables since conditioning the hazard on past processes allows the time-dependent explanatory variables to be thought of as fixed.

The following theorem outlines sufficient conditions for the uniform convergence of means of matrix products with variable inner dimensions.

*Theorem 4.1.* Let $\underline{X}_{hi}$ be an $rn_{hi}$ x 1 vector belonging to $\mathbb{R}^{rn_{hi}}$; $\underline{\alpha}$ and $\underline{\beta}$ be t x 1 parameter vectors belonging to $\Theta$, a compact subset of $\mathbb{R}^t$; and $A(\underline{\alpha},\underline{X}_{hi})$ and $B(\underline{\beta},\underline{X}_{hi})$ be $p \times n_{hi}$ and $n_{hi} \times q$ matrices, respectively, with elements that are uniformly bounded and continuous functions on $\Theta \times \mathbb{R}^{rn_{hi}}$. Suppose that assumptions A4.1 - A4.3, A4.5 and A4.6 hold. Then

$$\lim_{m\to\infty} k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} A(\underline{\alpha},\underline{X}_{hi})\ B(\underline{\beta},\underline{X}_{hi})$$

converges uniformly for all $\underline{\alpha}$ and $\underline{\beta} \in \Theta$.

*Proof.* Using the assumptions on A and B, assumption A4.6, and a multivariate extension of Theorem 2 in Jennrich (1969) derived for matrix products with common inner dimensions for all i (see Theorem 3.3.1 in Morel, 1987), for each h and $\delta$,

$$\lim_{m_{h\delta}\to\infty} m_{h\delta}^{-1} \sum_{i:n_{hi}=\delta} A(\underline{\alpha},\underline{X}_{hi})\ B(\underline{\beta},\underline{X}_{hi}) \equiv C_{h\delta}(\underline{\alpha},\underline{\beta})$$

converges uniformly for all $\underline{\alpha}$ and $\underline{\beta} \in \Theta$. Since the elements of A and B are continuous on a compact set, by Theorem 4.15 in Rudin (1976) the elements of $C_{h\delta}$ are also uniformly bounded. Hence for any value of h,

$$\lim_{m_h\to\infty} m_h^{-1} \sum_{i=1}^{m_h} A(\underline{\alpha},\underline{X}_{hi})\ B(\underline{\beta},\underline{X}_{hi})$$

$$= \lim_{m_h\to\infty} \sum_{\delta} (m_h^{-1} m_{h\delta})\ m_{h\delta}^{-1} \sum_{i:n_{hi}=\delta} A(\underline{\alpha},\underline{X}_{hi})\ B(\underline{\beta},\underline{X}_{hi})$$

$$= \sum_{\delta} \{ \lim_{m_h\to\infty} (m_h^{-1} m_{h\delta})$$

$$\times \lim_{m_{h\delta}\to\infty} m_{h\delta}^{-1} \sum_{i:n_{hi}=\delta} A(\underline{\alpha},\underline{X}_{hi})\; B(\underline{\beta},\underline{X}_{hi})\; )$$

(by assumption A4.3)

$$= \sum_{\delta} \Delta_{h\delta}\; C_{h\delta}(\underline{\alpha},\underline{\beta})$$

$$\equiv C_h(\underline{\alpha},\underline{\beta})$$

(by assumption A4.2), whose elements are uniformly bounded. Hence

$$\lim_{m\to\infty} k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} A(\underline{\alpha},\underline{X}_{hi})\; B(\underline{\beta},\underline{X}_{hi})$$

$$= k^{-1} \sum_{h=1}^{k} \lim_{m_h\to\infty} m_h^{-1} \sum_{i=1}^{m_h} A(\underline{\alpha},\underline{X}_{hi})\; B(\underline{\beta},\underline{X}_{hi})$$

$$= k^{-1} \sum_{h=1}^{k} C_h(\underline{\alpha},\underline{\beta}) \quad ,$$

(by assumption A4.5) which converges uniformly for all $\underline{\alpha}$ and $\underline{\beta} \in \Theta$. □

The following theorem is concerned with the convergence of weighted averages of differences. An assumption on the inverse of $V(\underline{\theta},\underline{X}_{hi})$ is required for its proof.

*Assumption A4.7.* The elements of $V^{-1}(\underline{\theta},\underline{X}_{hi})$ are uniformly bounded.

*Theorem 4.2.* Assume that model (4.1), assumptions A4.1 - A4.3 and A4.5 - A4.7 hold. Let $F(\underline{\gamma},\underline{X}_{hi})$ be a p x $n_{hi}$ matrix with elements consisting of uniformly bounded and continuous functions on $\Gamma \times \mathbb{R}^{rn_{hi}}$. Then

$$\lim_{m\to\infty} k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} F(\underline{\gamma},\underline{X}_{hi}) V^{-1}(\underline{\theta},\underline{X}_{hi})[\underline{Y}_{hi}-\underline{\pi}(\underline{\gamma},\underline{X}_{hi})] = \underline{0}$$

uniformly for all $\underline{\gamma} \in \Gamma$.

*Proof.* By a multivariate extension of Theorem 3 in Jennrich (1969) (see Theorem 3.3.2 in Morel, 1987) and assumptions A4.6 and A4.7, for each value of h and $\delta$,

$$\lim_{m_{h\delta}\to\infty} m_{h\delta}^{-1} \sum_{i:n_{hi}=\delta} F(\underline{\gamma},\underline{X}_{hi})\ V^{-1}(\underline{\theta},\underline{X}_{hi})\ [\underline{Y}_{hi} - \underline{\pi}(\underline{\gamma},\underline{X}_{hi})]$$

$$= \underline{0} \quad \text{a.s.}$$

uniformly for all $\underline{\gamma} \in \Gamma$. Hence

$$\lim_{m_h\to\infty} m_h^{-1} \sum_{i=1}^{m_h} F(\underline{\gamma},\underline{X}_{hi})\ V^{-1}(\underline{\theta},\underline{X}_{hi})\ [\underline{Y}_{hi} - \underline{\pi}(\underline{\gamma},\underline{X}_{hi})]$$

$$= \sum_{\delta} \lim_{m_h\to\infty} (m_h^{-1} m_{h\delta}) \lim_{m_{h\delta}\to\infty} m_{h\delta}^{-1} \sum_{i:n_{hi}=\delta} \{F(\underline{\gamma},\underline{X}_{hi})$$

$$\times V^{-1}(\underline{\theta},\underline{X}_{hi})\ [\underline{Y}_{hi} - \underline{\pi}(\underline{\gamma},\underline{X}_{hi})]\}$$

(by assumption A4.3)

$$= \sum_{\delta} \Delta_{h\delta}\ \underline{0} \quad \text{a.s.}$$

uniformly for all $\underline{\gamma} \in \Gamma$ (by assumption A4.2)

$$= \underline{0} \quad \text{a.s.}$$

uniformly for all $\underline{\gamma} \in \Gamma$. Hence

$$\lim_{m\to\infty} k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} F(\underline{\gamma},\underline{X}_{hi})\, V^{-1}(\underline{\theta},\underline{X}_{hi})\, [\underline{Y}_{hi} - \underline{\pi}(\underline{\gamma},\underline{X}_{hi})]$$

$$= k^{-1} \sum_{h=1}^{k} \lim_{m_h\to\infty} m_h^{-1} \sum_{i=1}^{m_h} F(\underline{\gamma},\underline{X}_{hi})\, V^{-1}(\underline{\theta},\underline{X}_{hi})$$

$$\times\, [\underline{Y}_{hi} - \underline{\pi}(\underline{\gamma},\underline{X}_{hi})]$$

(by assumption A4.5)

$$= \underline{0} \quad \text{a.s.}$$

uniformly for all $\underline{\gamma} \in \Gamma$. □

## 4. Nonlinear Least Squares Estimation of Mean Model Parameters

### 4.4.1. A Nonlinear Least Squares Estimator for $\underline{\gamma}$ When $V(\underline{\theta},\underline{X}_{hi})$ is Known

Consider the case where $V(\underline{\theta},\underline{X}_{hi})$, the covariance matrix of $\underline{Y}_{hi}$, is known for all h and i. The unknown parameter vector $\underline{\gamma}$ in model (4.1) can be estimated using a modified version of the traditional multivariate nonlinear least squares estimator. For m groups, define the nonlinear least squares estimator of $\underline{\gamma}$ to be the value of $\underline{\gamma}$ that minimizes

$$Q_m(\underline{\gamma}) = k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} \{ [\underline{Y}_{hi} - \underline{\pi}(\underline{\gamma}, \underline{X}_{hi})]' \, V^{-1}(\underline{\theta}, \underline{X}_{hi}) \times [\underline{Y}_{hi} - \underline{\pi}(\underline{\gamma}, \underline{X}_{hi})] \} \ . \qquad (4.2)$$

In practice, $V(\underline{\theta}, \underline{X}_{hi})$ is rarely known. In this case, least squares estimators based on the Gauss-Newton algorithm can be used to estimate $\underline{\gamma}$. Before discussing least squares estimators obtained from the Gauss-Newton algorithm, the following result will be proven for use in deriving the asymptotic distribution of Gauss-Newton estimators.

### 4.4.2. An Asymptotically Normal Pseudo Estimator

In this section, a pseudo estimator, $\ddot{\underline{\gamma}}_m$, is developed along with sufficient conditions for asymptotic normality as the number of groups, m, gets large. Let the true value of $\underline{\theta}$ be denoted $\underline{\theta}_o = (\underline{\gamma}_o', \underline{\alpha}_o')'$. As will be seen below, $\ddot{\underline{\gamma}}_m$ is not a true estimator since it depends on $\underline{\gamma}_o$ and thus cannot be calculated from the data.

An approximate expression for

$$\underline{e}_{hi} = \underline{Y}_{hi} - \underline{\pi}(\underline{\gamma}_o, \underline{X}_{hi})$$

can be derived by using a first order Taylor series approximation of $\underline{\pi}(\underline{\gamma}, \underline{X}_{hi})$ about $\underline{\gamma}_o$,

$$\underline{\pi}(\underline{\gamma}, \underline{X}_{hi}) = \underline{\pi}(\underline{\gamma}_o, \underline{X}_{hi}) + D(\underline{\gamma}_o, \underline{X}_{hi}) \; (\underline{\gamma} - \underline{\gamma}_o) + \underline{r}_{hi}(\underline{\gamma}_{hi}^{*}) \ ,$$

where

$$D(\underline{\gamma},\underline{X}) = \partial\underline{\pi}(\underline{\gamma},\underline{X})/\partial\underline{\gamma}' ,$$

$\underline{r}_{hi}(\underline{\gamma}^*_{hi})$ is the remainder term depending on the second derivative of $\underline{\pi}(\underline{\gamma}_o,\underline{X}_{hi})$ and $(\underline{\gamma} - \underline{\gamma}_o)$, and $\underline{\gamma}^*_{hi}$ lies on the line segment between $\underline{\gamma}$ and $\underline{\gamma}_o$ for each h and i. Thus

$$\underline{e}_{hi} = D(\underline{\gamma}_o,\underline{X}_{hi})(\underline{\gamma} - \underline{\gamma}_o) + [\underline{Y}_{hi} - \underline{\pi}(\underline{\gamma},\underline{X}_{hi}) + \underline{r}_{hi}(\underline{\gamma}^*_{hi})] .$$

Hence consider the linear model

$$\underline{e}_{hi} = D(\underline{\gamma}_o,\underline{X}_{hi})(\underline{\gamma} - \underline{\gamma}_o) + \underline{u}_{hi} , \qquad (4.3)$$

where the $\underline{u}_{hi}$ are independent random vectors with mean $\underline{0}$ and covariance matrix $V(\underline{\theta}_o,\underline{X}_{hi})$. This model implies that $\underline{\gamma} - \underline{\gamma}_o = \underline{0}$; i.e., the estimated value for $\underline{\gamma}$ is an estimate of $\underline{\gamma}_o$. Although $\mathrm{Var}(\underline{u}_{hi})$ is a function of $\underline{\theta}_o$, the development in this section is concerned only with the estimation of $\underline{\gamma}_o$, and it will be convenient to think of $\underline{\alpha}_o$ as a nuisance parameter and $V(\underline{\theta}_o,\underline{X}_{hi})$ as a function of $\underline{\gamma}_o$.

An estimator of $\underline{\gamma}_o$, $\ddot{\underline{\gamma}}_m$, can be constructed as the value of $\underline{\gamma}$ that minimizes

$$\ddot{Q}_m(\underline{\gamma}) = k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} \{ [\underline{e}_{hi} - D(\underline{\gamma}_o,\underline{X}_{hi})(\underline{\gamma} - \underline{\gamma}_o)]' \\ \times V^{-1}(\underline{\theta}_o,\underline{X}_{hi}) [\underline{e}_{hi} - D(\underline{\gamma}_o,\underline{X}_{hi})(\underline{\gamma}-\underline{\gamma}_o)] \}$$

over $\underline{\gamma} \in \Gamma$.

A few regularity conditions are required in order to prove the asymptotic normality of $\ddot{\underline{\gamma}}_m$.

*Assumption A4.8.* $\underline{\gamma}_o$ belongs to the interior of $\Gamma$.

Consider

$$W_m(\underline{\gamma}) = k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} D(\underline{\gamma},\underline{X}_{hi})' \; V^{-1}(\underline{\theta}_o,\underline{X}_{hi}) \; D(\underline{\gamma},\underline{X}_{hi}) \; . \tag{4.4}$$

Under the assumptions of model (4.1), the elements of $D(\underline{\gamma},\underline{X}_{hi})$ are continuous functions on a compact set and hence are uniformly bounded. Also, by assumption A4.7, the elements of $V^{-1}(\underline{\theta},\underline{X}_{hi})$ are uniformly bounded. So by Theorem 4.1, $W_m(\underline{\gamma})$ converges uniformly for all $\underline{\gamma} \in \Gamma$ to a limit, say $W(\underline{\gamma})$. One additional assumption is needed concerning the nonsingularity of $W(\underline{\gamma})$.

*Assumption A4.9.* There exists a neighborhood of $\underline{\gamma}_o$, denoted $N(\underline{\gamma}_o)$, such that $W(\underline{\gamma})$ is nonsingular for all $\underline{\gamma} \in N(\underline{\gamma}_o)$.

We will now prove the asymptotic normality of the pseudo estimator $\ddot{\underline{\gamma}}_m$.

*Theorem 4.3.* Suppose model (4.3) and assumptions A4.1 - A4.9 hold. Then

$$m^{1/2}(\ddot{\underline{\gamma}}_m - \underline{\gamma}_o) \xrightarrow{L} N_s(\underline{0}, [k\, W(\underline{\gamma}_o)]^{-1})$$

as $m \to \infty$ .

*Proof.* We begin by finding a useful expression for $\ddot{\underline{\gamma}}_m - \underline{\gamma}_o$. A multivariate central limit theorem is then applied to this expression to derive its limiting distribution.

First consider $\partial\ddot{Q}_m(\underline{\gamma})/\partial\underline{\gamma}'$. Observe that the derivative of the hi-th term of $\ddot{Q}_m(\underline{\gamma})$ with respect to $\underline{\gamma}'$ is

$$\partial\{\ [\underline{e}_{hi} - D(\underline{\gamma}_o,\underline{X}_{hi})(\underline{\gamma} - \underline{\gamma}_o)]'\ V^{-1}(\underline{\theta}_o,\underline{X}_{hi})$$

$$\times\ [\underline{e}_{hi} - D(\underline{\gamma}_o,\underline{X}_{hi})(\underline{\gamma} - \underline{\gamma}_o)]\ \}/\partial\underline{\gamma}'$$

$$= -\ 2\ D(\underline{\gamma}_o,\underline{X}_{hi})'\ V^{-1}(\underline{\theta}_o,\underline{X}_{hi})\ \underline{e}_{hi}$$

$$+\ 2\ D(\underline{\gamma}_o,\underline{X}_{hi})'\ V^{-1}(\underline{\theta}_o,\underline{X}_{hi})\ D(\underline{\gamma}_o,\underline{X}_{hi})\ (\underline{\gamma}_o-\underline{\gamma})\ .$$

Thus, setting $\partial\ddot{Q}_m(\underline{\gamma})/\partial\underline{\gamma}'$ to $\underline{0}$ yields

$$\underline{0} = k^{-1}\sum_{h=1}^{k}\ m_h^{-1}\sum_{i=1}^{m_h} D(\underline{\gamma}_o,\underline{X}_{hi})'\ V^{-1}(\underline{\theta}_o,\underline{X}_{hi})\ \underline{e}_{hi}$$

$$-\ W(\underline{\gamma}_o)(\underline{\gamma} - \underline{\gamma}_o)\ .$$

Note that by definition $\ddot{\underline{\gamma}}_m$ is a solution to this equation. Further, assumption A4.9 implies that $W_m(\underline{\gamma}_o)$ is nonsingular for large m. So for sufficiently large m,

$$\ddot{\underline{\gamma}}_m - \underline{\gamma}_o = W_m^{-1}(\underline{\gamma}_o)$$

$$\times\ k^{-1}\sum_{h=1}^{k}\ m_h^{-1}\sum_{i=1}^{m_h} D(\underline{\gamma}_o,\underline{X}_{hi})'V^{-1}(\underline{\theta}_o,\underline{X}_{hi})\underline{e}_{hi}\ . \tag{4.5}$$

Consider

$$m_h^{-1} \sum_{i=1}^{m_h} D(\underline{\gamma}_o, \underline{X}_{hi})' \ V^{-1}(\underline{\theta}_o, \underline{X}_{hi}) \ \underline{e}_{hi} \qquad (\dagger)$$

for any h. A multivariate central limit theorem in Rao (1973, p. 147) can be used to show that (†) is asymptotically normal as $m \to \infty$. This theorem requires that two conditions be met. First note that

$$\underline{Z}_{hi} = D(\underline{\gamma}_o, \underline{X}_{hi})' \ V^{-1}(\underline{\theta}_o, \underline{X}_{hi}) \ \underline{e}_{hi}$$

are independent for all i with mean $\underline{0}$ and covariance matrix

$$D(\underline{\gamma}_o, \underline{X}_{hi})' \ V^{-1}(\underline{\theta}_o, \underline{X}_{hi}) \ D(\underline{\gamma}_o, \underline{X}_{hi}) \ .$$

Then, since the elements of $\mathrm{Var}\{\underline{Z}_{hi}\}$ are uniformly bounded for all h and i,

$$\lim_{m\to\infty} m_h^{-1} \sum_{i=1}^{m_h} \mathrm{Var}\{\underline{Z}_{hi}\} \equiv W_h(\underline{\gamma}_o)$$

exists and is a nonzero matrix, and the first condition is satisfied. For second condition required is that

$$\lim_{m\to\infty} m_h^{-1} \sum_{i=1}^{m_h} \int_{\|\underline{z}\|>\varepsilon\sqrt{m}} \|\underline{z}\|^{\,2} dG_{hi}(\underline{z}) = \underline{0} \ ,$$

where $\|\cdot\|$ denotes the euclidean norm of a vector and $G_{hi}$ is the distribution function of $\underline{Z}_{hi}$. Since the elements of $\underline{Z}_{hi}$ are uniformly bounded, there exists a constant B such that for all m,

$$\|\underline{Z}_{hi}\|^2 \le B$$

for all h and i. Then for any $\varepsilon > 0$,

$$0 \le \int_{\|\underline{z}\|>\varepsilon\sqrt{m}} \|\underline{z}\|^2 \, dG_{hi}(\underline{z})$$

$$\le B \int_{\|\underline{z}\|>\varepsilon\sqrt{m}} dG_{hi}(\underline{z})$$

$$= B \Pr\{\|\underline{Z}_{hi}\| > \varepsilon\sqrt{m}\}$$

$$\le (\varepsilon^2 m)^{-1} B \, E\{\|\underline{Z}_{hi}\|^2\}$$

(by Chebychev's inequality)

$$\le (\varepsilon^2 m)^{-1} B^2 .$$

Hence,

$$0 \le \lim_{m\to\infty} m_h^{-1} \sum_{i=1}^{m_h} \int_{\|\underline{z}\|>\varepsilon\sqrt{m}} \|\underline{z}\|^2 dG_{hi}(\underline{z})$$

$$\le \lim_{m\to\infty} (m_h^{-1} m) \lim_{m\to\infty} (\varepsilon^2 m)^{-1} B^2$$

$$= 0$$

(by assumption A4.4).

Since the conditions necessary for application of the multivariate central limit theorem are met, as $m\to\infty$

$$m^{1/2}[m_h^{-1} \sum_{i=1}^{m_h} D(\underline{\gamma}_o,\underline{X}_{hi})' \, V^{-1}(\underline{\theta}_o,\underline{X}_{hi}) \, \underline{e}_{hi}]$$

$$\xrightarrow{L} N_s[\underline{0}, W_h(\underline{\gamma}_o)] ,$$

implying that as $m\to\infty$

$$m^{1/2}[k^{-1}\sum_{h=1}^{k} m_h^{-1}\sum_{i=1}^{m_h} D(\underline{\gamma}_o,\underline{X}_{hi})'\; V^{-1}(\underline{\theta}_o,\underline{X}_{hi})\; \underline{e}_{hi}]$$

$$\xrightarrow{L} N_s[\underline{0},\; k^{-1}\; W(\underline{\gamma}_o)]$$

Finally, since $W_m$ converges uniformly to W and is nonsingular for sufficiently large m, and the elements of $W_m^{-1}$ are continuous functions of $W_m$,

$$\lim_{m\to\infty} W_m^{-1}(\underline{\gamma}) = W^{-1}(\underline{\gamma}) \;.$$

So by Corollary 5.2.6.2 in Fuller (1976),

$$m^{1/2}(\tilde{\underline{\gamma}}_m - \underline{\gamma}_o) \xrightarrow{L} N_s(\underline{0},\; [k\; W(\underline{\gamma}_o)]^{-1}) \;.$$

as $m\to\infty$. □

### 4.4.3. Nonlinear Least Squares Estimation for $\underline{\gamma}$ When $V(\underline{\theta},\underline{X}_{hi})$ is Unknown

In practice, the variance of $\underline{Y}_{hi}$ is rarely known and is related to the mean of $\underline{Y}_{hi}$ in addition to other unknown parameters, $\underline{\alpha}$. To address this situation, consider the use of multivariate nonlinear least squares estimators based on the Gauss-Newton algorithm (Gallant, 1987). These estimators are constructed by approximating the model for $\underline{Y}_{hi}$ using a first order Taylor series approximation about $\underline{\gamma}_o$ for the nonlinear mean function. Model (4.1) is approximated by

$$\underline{Y}_{hi} = \underline{\pi}(\underline{\gamma}_o, \underline{X}_{hi}) + D(\underline{\gamma}_o, \underline{X}_{hi})'(\underline{\gamma} - \underline{\gamma}_o) + \underline{u}_{hi} , \tag{4.6}$$

where the $\underline{u}_{hi}$ are independent random vectors with mean $\underline{0}$ and covariance matrix $V(\underline{\theta}_o, \underline{X}_{hi})$. The derivation for this model is the same as that for the model described in equation (4.3).

Obtaining parameter estimates using Gauss-Newton estimation involves an iterative procedure that adjusts the previous estimate at each step. The updating adjustment for the parameter estimate from the previous iteration is derived from (4.6) by noting that

$$(\underline{\gamma} - \underline{\gamma}_o) \cong [D(\underline{\gamma}_o, \underline{X}_{hi})' \; V^{-1}(\underline{\theta}_o, \underline{X}_{hi}) \; D(\underline{\gamma}_o, \underline{X}_{hi})]^{-1}$$
$$\times \; D(\underline{\gamma}_o, \underline{X}_{hi})' V^{-1}(\underline{\theta}_o, \underline{X}_{hi}) [\underline{Y}_{hi} - \underline{\pi}(\underline{\gamma}_o, \underline{X}_{hi})] \; .$$

Hence the estimated value of $\underline{\gamma}_o$ from the c-th iteration can be obtained from

$$\hat{\underline{\gamma}}_m^{(c)} = \hat{\underline{\gamma}}_m^{(c-1)}$$
$$+ \; W_m^{-1}(\hat{\underline{\theta}}^{(c-1)}) \; k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} \{ \; D(\hat{\underline{\gamma}}_m^{(c-1)}, \underline{X}_{hi})'$$
$$\times \; V^{-1}(\hat{\underline{\theta}}^{(c-1)}, \underline{X}_{hi}) \; [\underline{Y}_{hi} - \underline{\pi}(\hat{\underline{\gamma}}_m^{(c-1)}, \underline{X}_{hi})] \; \},$$

(4.7)

where

$$\hat{\underline{\theta}}^{(c-1)} = (\hat{\underline{\gamma}}_m^{(c-1)\prime}, \; \hat{\underline{\alpha}}^{(0)\prime})' \; ,$$

$$W_m(\underline{\theta}) = k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} D(\underline{\gamma},\underline{X}_{hi})' \, V^{-1}(\underline{\theta},\underline{X}_{hi}) \, D(\underline{\gamma},\underline{X}_{hi}) \; , \tag{4.8}$$

and $\hat{\underline{\alpha}}^{(0)}$ is the initial estimate of $\underline{\alpha}$ (alternatively, $\hat{\underline{\alpha}}^{(0)}$ can be updated at each step). The estimator defined by c Gauss-Newton iterations is often referred to as the c-step Gauss-Newton estimator of $\underline{\gamma}_o$. Note that (4.7) is of the same form as the pseudo estimator defined in (4.5). If consistent estimators, $\hat{\underline{\theta}}^{(0)}$, of the true mean and correlation parameters, $\underline{\theta}_o$, are used to initiate the iterative procedure, $\hat{\underline{\gamma}}_m^{(c)}$ can be shown to be asymptotically normal.

The asymptotic distribution of the one-step Gauss-Newton estimator [equation (4.7) with c = 1] is derived in Section 4.4.5 for data collected from independent groups with correlations among responses for individuals within groups. The asymptotic distribution of the general c-step Gauss-Newton estimator is then developed in Section 4.4.6 from the properties of the one-step Gauss-Newton estimator.

#### 4.4.5. The One-step Gauss-Newton Estimator

Assume that there exist consistent estimators of $\underline{\alpha}_o$ and $\underline{\gamma}_o$, denoted $\hat{\underline{\theta}}^{(0)} = (\hat{\underline{\alpha}}^{(0)\prime}, \hat{\underline{\gamma}}^{(0)\prime})'$, such that for some sequence of constants $\{a_m\}$ with $\lim_{m\to\infty} a_m = 0$,

$\hat{\underline{\theta}}^{(0)} = \underline{\theta}_o + o_p(a_m)$ .

To prove the asymptotic normality of $\hat{\underline{\gamma}}_m^{(1)}$, three additional assumptions are required.

*Assumption A4.10.* The elements of $V^{-1}(\underline{\theta},\underline{X}_{hi})$ are continuous functions of $\underline{\theta} \in \Theta$, with continuous first and second derivatives with respect to $\underline{\theta}$.

*Assumption A4.11.* $\underline{\alpha}_o$ belongs to the interior of $\Phi$.

Recall $W_m(\underline{\theta})$ as defined by equation (4.8). Since $D(\underline{\gamma},\underline{X})$ and $V^{-1}(\underline{\theta},\underline{X})$ are continuous on compact sets,

$$\lim_{m\to\infty} W_m(\underline{\theta}) = W(\underline{\theta}) \quad \text{a.s.}$$

uniformly for all $\underline{\theta} \in \Theta$. The last assumption required is an extension of assumption A4.9.

*Assumption A4.12.* For some neighborhood of $\underline{\theta}_o$, $N(\underline{\theta}_o) = N(\underline{\alpha}_o) \times N(\underline{\gamma}_o)$, $W(\underline{\theta})$ is nonsingular for all $\underline{\theta} \in N(\underline{\theta}_o)$.

To begin with, an asymptotically useful expression for $\hat{\underline{\gamma}}_m^{(1)} - \underline{\gamma}_o$ is needed. The following theorem provides this expression.

*Theorem 4.4.* Assume that model (4.1) holds, and that assumptions A4.1 - A4.8 and A4.10 - A4.12 hold. Let $\hat{\underline{\theta}}^{(0)}$ be a consistent estimator of $\underline{\theta}_o$ such that for some sequence $\{a_m\}$ with $\lim_{m\to\infty} a_m = 0$ ,

$$\hat{\underline{\theta}}^{(0)} = \underline{\theta}_o + O_p(a_m).$$

Let $\hat{\gamma}_m^{(1)}$ be the one-step Gauss-Newton estimator defined in equation (4.7). Then

$$\hat{\gamma}_m^{(1)} - \gamma_o$$

$$= W_m^{-1}(\underline{\theta}_o)\ k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} D(\gamma_o, \underline{X}_{hi})'\ V^{-1}(\underline{\theta}_o, \underline{X}_{hi})\ \underline{e}_{hi}$$

$$+ O_p[\max(m^{-1/2}a_m,\ a_m^2)]\ .$$

*Proof.* An expression for $\hat{\gamma}_m^{(1)} - \gamma_o$ will be derived, and then its asymptotic behavior will be examined.

Expanding $\underline{\pi}(\gamma_o, \underline{X}_{hi})$ in a Taylor series about $\hat{\gamma}^{(0)}$ yields

$$\underline{\pi}(\gamma_o, \underline{X}_{hi}) - \underline{\pi}(\hat{\gamma}^{(0)}, \underline{X}_{hi})$$

$$= D(\hat{\gamma}^{(0)}, \underline{X}_{hi})\ (\gamma_o - \hat{\gamma}^{(0)}) + \underline{r}(\hat{\gamma}^{(0)}, \underline{X}_{hi})$$

where

$$\underline{r}(\hat{\gamma}^{(0)}, \underline{X}_{hi})$$

$$= 2^{-1}\Big[(\gamma_o - \hat{\gamma}^{(0)})'\Big\{\partial^2\pi_1(\gamma_{hi}^*, \underline{X}_{hi})/\partial\gamma_o\partial\gamma_o'\Big\}(\gamma_o - \hat{\gamma}^{(0)}),$$

$$\ldots,\ (\underline{\gamma}_o-\hat{\underline{\gamma}}^{(0)})'\left\{\partial^2\pi_{n_{hi}}(\underline{\gamma}^*_{hi},\underline{X}_{hi})/\partial\underline{\gamma}_o\partial\underline{\gamma}_o'\right\}(\underline{\gamma}_o-\hat{\underline{\gamma}}^{(0)})\Big]'$$

and where the elements of $\underline{\gamma}^*_{hi}$ lie between the corresponding elements of $\underline{\gamma}_o$ and $\hat{\underline{\gamma}}^{(0)}$ for all h and i. Using this expression, under model (4.1),

$$\begin{aligned}
&\underline{Y}_{hi} - \underline{\pi}(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi}) \\
&\quad = \underline{\pi}(\underline{\gamma}_o,\underline{X}_{hi}) + \underline{e}_{hi} - \underline{\pi}(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi}) \\
&\quad = D(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})(\underline{\gamma}_o-\hat{\underline{\gamma}}^{(0)}) + \underline{r}(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi}) + \underline{e}_{hi}\ , \qquad (\dagger)
\end{aligned}$$

which implies that for sufficiently large m

$$\begin{aligned}
&(\hat{\underline{\gamma}}^{(0)} - \underline{\gamma}_o) \\
&\quad = -\Big[\ k^{-1}\sum_{h=1}^{k} m_h^{-1}\sum_{i=1}^{m_h} D(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})'\ V^{-1}(\hat{\underline{\theta}}^{(0)},\underline{X}_{hi}) \\
&\qquad\qquad \times D(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})\ \Big]^{-1} \\
&\qquad \times k^{-1}\sum_{h=1}^{k} m_h^{-1}\sum_{i=1}^{m_h} D(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})'\ V^{-1}(\hat{\underline{\theta}}^{(0)},\underline{X}_{hi}) \\
&\qquad\qquad \times \{\ [\underline{Y}_{hi} - \underline{\pi}(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})] - \underline{r}(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi}) \\
&\qquad\qquad\qquad - \underline{e}_{hi}\ \}\ . \\
&\quad = -\ W_m^{-1}(\hat{\underline{\theta}}^{(0)})\ k^{-1}\sum_{h=1}^{k} m_h^{-1}\sum_{i=1}^{m_h}\{\ D(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})' \\
&\qquad\qquad \times V^{-1}(\hat{\underline{\theta}}^{(0)},\underline{X}_{hi})\ [\underline{Y}_{hi} - \underline{\pi}(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})]\ \} \\
&\qquad + W_m^{-1}(\hat{\underline{\theta}}^{(0)})\ k^{-1}\sum_{h=1}^{k} m_h^{-1}\sum_{i=1}^{m_h}\{\ D(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})'
\end{aligned}$$

$$\times\ V^{-1}(\hat{\underline{\theta}}^{(0)},\underline{X}_{hi})\ \underline{r}(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})\ \}$$

$$+\ W_m^{-1}(\hat{\underline{\theta}}^{(0)})\ k^{-1}\sum_{h=1}^{k} m_h^{-1}\sum_{i=1}^{m_h} D(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})'$$

$$\times\ V^{-1}(\hat{\underline{\theta}}^{(0)},\underline{X}_{hi})\ \underline{e}_{hi}\ .$$

From this expression for $\hat{\underline{\gamma}}^{(0)}-\underline{\gamma}_o$ and the definition for $\hat{\underline{\gamma}}_m^{(1)}$ in equation (4.7) with c = 1, $\hat{\underline{\gamma}}_m^{(1)}-\underline{\gamma}_o$ can be rewritten as

$$\hat{\underline{\gamma}}_m^{(1)} - \underline{\gamma}_o$$

$$= \hat{\underline{\gamma}}^{(0)} - \underline{\gamma}_o$$

$$+\ W_m\ (\hat{\underline{\theta}}^{(0)})\ k^{-1}\sum_{h=1} m_h^{-1}\sum_{i=1}^{h}\ \{\ D(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})'$$

$$\times\ V^{-1}(\hat{\underline{\theta}}^{(0)},\underline{X}_{hi})\ [\underline{Y}_{hi} - \underline{\pi}(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})]\ \}$$

$$= W_m^{-1}(\hat{\underline{\theta}}^{(0)})\ k^{-1}\sum_{h=1}^{k} m_h^{-1}\sum_{i=1}^{m_h}\ \{\ D(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})'$$

$$\times\ V^{-1}(\hat{\underline{\theta}}^{(0)},\underline{X}_{hi})\ \underline{r}(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})\ \}$$

$$+\ W_m^{-1}(\hat{\underline{\theta}}^{(0)})\ k^{-1}\sum_{h=1}^{k} m_h^{-1}\sum_{i=1}^{m_h}\ \{\ D(\hat{\underline{\gamma}}^{(0)},\underline{X}_{hi})'$$

$$\times\ V^{-1}(\hat{\underline{\theta}}^{(0)},\underline{X}_{hi})\ \underline{e}_{hi}\ \}$$

= term 1 + term 2 .

Consider the asymptotic behavior of term 1, which will be shown to be $O_p(a_m^2)$. Let the j-th row of

$$k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} D(\hat{\underline{\gamma}}^{(0)}, \underline{X}_{hi}) V^{-1}(\hat{\underline{\theta}}^{(0)}, \underline{X}_{hi}) \underline{r}(\hat{\underline{\gamma}}^{(0)}, \underline{X}_{hi}) \qquad (\ddagger)$$

be denoted $z_{1j}(\hat{\underline{\theta}}^{(0)})$, and let $v^{ab}(\hat{\underline{\theta}}^{(0)}, \underline{X}_{hi})$ represent element (a,b) in $V^{-1}(\hat{\underline{\theta}}^{(0)}, \underline{X}_{hi})$. Then

$$z_{1j}(\hat{\underline{\theta}}^{(0)})$$

$$= k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} \sum_{a=1}^{n_{hi}} \sum_{b=1}^{n_{hi}} \{ \partial\pi_a(\hat{\underline{\gamma}}^{(0)}, \underline{X}_{hi})/\partial\gamma_j$$

$$\times\ v^{ab}(\hat{\underline{\theta}}^{(0)}, \underline{X}_{hi})\ [\ 2^{-1}(\underline{\gamma}_o - \hat{\underline{\gamma}}^{(0)})'$$

$$\times\ \partial^2\pi_b(\underline{\gamma}^*_{hi}, \underline{X}_{hi})/\partial\underline{\gamma}\partial\underline{\gamma}'\ (\underline{\gamma}_o - \hat{\underline{\gamma}}^{(0)})\ ]\ \}$$

$$= k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} \sum_{a=1}^{n_{hi}} \sum_{b=1}^{n_{hi}} \sum_{c=1}^{s} \sum_{d=1}^{s} \{ \partial\pi_a(\hat{\underline{\gamma}}^{(0)}, \underline{X}_{hi})/\partial\gamma_j$$

$$\times\ v^{ab}(\hat{\underline{\theta}}^{(0)}, \underline{X}_{hi})\ 2^{-1}\partial^2\pi_b(\underline{\gamma}^*_{hi}, \underline{X}_{hi})/\partial\gamma_c\partial\gamma_d$$

$$\times\ (\gamma_{oc} - \hat{\gamma}_c^{(0)})(\gamma_{od} - \hat{\gamma}_d^{(0)})\ \}$$

$$= \sum_{\gamma} k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i:n_{hi}=\gamma} \sum_{a=1}^{\gamma} \sum_{b=1}^{\gamma} \sum_{c=1}^{s} \sum_{d=1}^{s}$$

$$\times\ \{\partial\pi_a(\hat{\underline{\gamma}}^{(0)}, \underline{X}_{hi})/\partial\gamma_j$$

$$\times\ v^{ab}(\hat{\underline{\theta}}^{(0)}, \underline{X}_{hi})\ 2^{-1}\partial^2\pi_b(\underline{\gamma}^*_{hi}, \underline{X}_{hi})/\partial\gamma_c\partial\gamma_d$$

$$\times\ (\gamma_{oc} - \hat{\gamma}_c^{(0)})(\gamma_{od} - \hat{\gamma}_d^{(0)})\ \}$$

$$= \sum_{\gamma} \sum_{a=1}^{\gamma} \sum_{b=1}^{\gamma} \sum_{c=1}^{s} \sum_{d=1}^{s}$$

$$\times \left\{ \left[ k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i:n_{hi}=\gamma} \left( \partial\pi_a(\hat{\underline{\gamma}}^{(0)}, \underline{X}_{hi})/\partial\gamma_j \right.\right.\right.$$

$$\left.\left. \times\, v^{ab}(\hat{\underline{\theta}}^{(0)}, \underline{X}_{hi})\; 2^{-1}\partial^2\pi_b(\underline{\gamma}^*_{hi}, \underline{X}_{hi})/\partial\gamma_c\partial\gamma_d \right) \right]$$

$$\left. \times\, (\gamma_{oc} - \hat{\gamma}_c^{(0)})\; (\gamma_{od} - \hat{\gamma}_d^{(0)}) \right\} .$$

Now since $\partial\pi_a(\hat{\underline{\gamma}}^{(0)}, \underline{X}_{hi})/\partial\gamma_j$ and $v^{ab}(\hat{\underline{\theta}}^{(0)}, \underline{X}_{hi})$ are uniformly bounded and continuous, as $m \to \infty$

$$k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i:n_{hi}=\gamma} [\, \partial\pi_a(\underline{\gamma}, \underline{X}_{hi})/\partial\gamma_j \; v^{ab}(\underline{\theta}, \underline{X}_{hi}) \,]^2$$

converges uniformly for all $\underline{\theta} \in \Phi \times \Gamma$ to say $L_{1jabcd}(\underline{\theta})$, which by Theorem 4.15 in Rudin (1976) is also bounded. Likewise,

$$k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i:n_{hi}=\gamma} [\, \partial^2\pi_b(\underline{\gamma}, \underline{X}_{hi})/\partial\gamma_c\partial\gamma_d \,]^2$$

converges uniformly for all $\underline{\theta} \in \Phi \times \Gamma$ to bounded limit $L_{2jabcd}(\underline{\theta})$. Hence, for all $\varepsilon_1 > 0$, there exists an integer $M_{\varepsilon 1}$ such that for all $m > M_{\varepsilon 1}$

$$\Big| \, k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i:n_{hi}=\gamma} [\, \partial\pi_a(\underline{\gamma}, \underline{X}_{hi})/\partial\gamma_j \; v^{ab}(\underline{\theta}, \underline{X}_{hi}) \,]^2$$

$$- L_{1jabcd}(\underline{\theta}) \, \Big| < \varepsilon_1$$

and

$$\Big| \, k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i:n_{hi}=\gamma} [\, \partial^2\pi_b(\underline{\gamma}, \underline{X}_{hi})/\partial\gamma_c\partial\gamma_d \,]^2$$

$$- L_{2jabcd}(\underline{\theta}) \mid < \varepsilon_1$$

for all $\underline{\theta} \in \Phi \times \Gamma$. Further,

$$\hat{\underline{\theta}}^{(0)} - \underline{\theta}_o = O_p(a_m)$$

implies that for all $\varepsilon_2 > 0$, there exists an integer $M_{\varepsilon 2}$ such that for all $m > M_{\varepsilon 2}$

$$\Pr\{ \hat{\underline{\theta}}^{(0)} \in N(\underline{\theta}_o) \} > 1 - \varepsilon_2 .$$

Hence for $m > \max(M_{\varepsilon 1}, M_{\varepsilon 2},)$ and $\hat{\underline{\theta}}^{(0)} \in N(\underline{\theta}_o)$, by the Cauchy-Schwartz inequality,

$$k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i:n_{hi}=\gamma} \{ \partial\pi_a(\hat{\underline{\gamma}}^{(0)}, \underline{X}_{hi})/\partial\gamma_j \; v^{ab}(\hat{\underline{\theta}}^{(0)}, \underline{X}_{hi})$$

$$\times\, 2^{-1}\partial^2\pi_b(\underline{\gamma}^{*}_{hi}, \underline{X}_{hi})/\partial\gamma_c\partial\gamma_d \}$$

$$\le 2^{-1}\left[k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i:n_{hi}=\gamma} [\partial\pi_a(\underline{\gamma}, \underline{X}_{hi})/\partial\gamma_j \; v^{ab}(\underline{\theta}, \underline{X}_{hi})]^2\right.$$

$$\left.\times\, k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i:n_{hi}=\gamma} [\partial^2\pi_b(\underline{\gamma}, \underline{X}_{hi})/\partial\gamma_c\partial\gamma_d]^2\right]^{1/2}$$

$$= 2^{-1} [\, O_p(1)\; O_p(1)\, ]^{1/2}$$

$$= O_p(1) .$$

Further, because $(\gamma_{oc} - \hat{\gamma}_c^{(0)})$ and $(\gamma_{od} - \hat{\gamma}_d^{(0)})$ are $O_p(a_m)$, (‡) is $O_p(a_m^2)$. Finally, since each element of $W_m(\underline{\theta})$ is a continuous function on a compact space (i.e., is uniformly continuous),

$$W_m(\hat{\underline{\theta}}^{(0)}) = W(\underline{\theta}_o) + o_p(a_m) \ .$$

So for m sufficiently large such that $W_m$ is nonsingular,

$$W_m^{-1}(\hat{\underline{\theta}}^{(0)}) = W_m^{-1}(\underline{\theta}_o) + o_p(1) \ ,$$

implying that

$$W_m^{-1}(\hat{\underline{\theta}}^{(0)}) = O_p(1) \ .$$

Therefore, term 1 is $o_p(a_m^2)$.

Now consider the j-th row of

$$k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} D(\hat{\underline{\gamma}}^{(0)}, \underline{X}_{hi})' \ V^{-1}(\hat{\underline{\theta}}^{(0)}, \underline{X}_{hi}) \ \underline{e}_{hi}$$

in term 2, denoted $z_{2j}(\hat{\underline{\theta}}^{(0)})$. By derivations similar to those for $z_{1j}(\hat{\underline{\theta}}^{(0)})$,

$$z_{2j}(\hat{\underline{\theta}}^{(0)})$$

$$= \sum_{\gamma} \sum_{a=1}^{\gamma} \sum_{b=1}^{\gamma} \left[ k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} \{ \partial\pi_a(\hat{\underline{\gamma}}^{(0)}, \underline{X}_{hi})/\partial\gamma_j \right.$$

$$\left. \times \ v^{ab}(\hat{\underline{\theta}}^{(0)}, \underline{X}_{hi}) \ e_{hib} \ \} \right]$$

$$\equiv \sum_{\gamma} \sum_{a=1}^{\gamma} \sum_{b=1}^{\gamma} z_{2jabm}(\hat{\underline{\theta}}^{(0)}) \ .$$

The asymptotic behavior of $z_{2jabm}(\hat{\underline{\theta}}^{(0)})$ can be examined by expanding $z_{2jabm}(\hat{\underline{\theta}}^{(0)})$ in a Taylor series about $\underline{\theta}_o$. For $\underline{\theta}_{cdhi}^{**}$ whose elements lie between the corresponding elements of $\hat{\underline{\theta}}^{(0)}$ and $\underline{\theta}_o$ for all c, d, h and i,

$$z_{2jabm}(\hat{\underline{\theta}}^{(0)})$$

$$= z_{2jabm}(\underline{\theta}_o)$$

$$+ \sum_{c=1}^{s+u} \partial z_{2jabm}(\underline{\theta}_o)/\partial\theta_c \; (\theta_c \quad - \theta_{oc})$$

$$+ 2^{-1} \sum_{c=1}^{s+u} \sum_{d=1}^{s+u} \{ \; \partial^2 z_{2jabm}(\underline{\theta}_{cd}^{**})/\partial\theta_c\partial\theta_d$$

$$\times \; (\hat{\theta}_c^{(0)} - \theta_{oc}) \; (\hat{\theta}_d^{(0)} - \theta_{od}) \; \}$$

where

$$\underline{\theta}_{cd}^{**} = (\underline{\theta}_{cd11}^{**\prime}, \; \ldots, \; \underline{\theta}_{cd1m_1}^{**\prime}, \; \underline{\theta}_{cd21}^{**\prime}, \; \ldots, \; \underline{\theta}_{cdkm_k}^{**\prime})' \; ,$$

$$\partial z_{2jabm}(\underline{\theta}_o)/\partial\theta_c$$

$$= k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} \{ \; [\partial^2\pi_a(\underline{\gamma}_o,\underline{X}_{hi})/\partial\gamma_c\partial\gamma_j \; v^{ab}(\underline{\theta}_o,\underline{X}_{hi})$$

$$+ \; \partial\pi_a(\underline{\gamma}_o,\underline{\gamma}_{hi})/\partial\gamma_j \; \partial v^{ab}(\underline{\theta}_o,\underline{X}_{hi})/\partial\theta_c] \; e_{hia} \; \}$$

$$\equiv k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} p_{jabc}(\underline{\theta}_o,\underline{X}_{hi}) \; e_{hia} \; ,$$

and

$$\partial^2 z_{2jabm}(\underline{\theta}^{**})/\partial\theta_d\partial\theta_c$$

$$= k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} ( \; [ \; \partial^3\pi_a(\underline{\gamma}_{cdhi}^{**},\underline{X}_{hi})/\partial\theta_d\partial\theta_c\partial\gamma_j$$

$$\times\ v^{ab}(\underline{\theta}^{**}_{cdhi},\underline{X}_{hi})$$

$$+\ \partial^2\pi_a(\underline{\gamma}^{**}_{cdhi},\underline{X}_{hi})/\partial\theta_c\partial\gamma_j\ \partial v^{ab}(\underline{\theta}^{**}_{cdhi},\underline{X}_{hi})/\partial\theta_d$$

$$+\ \partial^2\pi_a(\underline{\gamma}^{**}_{cdhi},\underline{X}_{hi})/\partial\theta_d\partial\gamma_j\ \partial v^{ab}(\underline{\theta}^{**}_{cdhi},\underline{X}_{hi})/\partial\theta_c$$

$$+\ \partial\pi_a(\underline{\gamma}^{**}_{cdhi},\underline{X}_{hi})/\partial\gamma_j\ \partial^2 v^{ab}(\underline{\theta}^{**}_{cdhi},\underline{X}_{hi})/\partial\theta_d\partial\theta_c]$$

$$\times\ e_{hia}\ \}$$

$$\equiv k^{-1}\sum_{h=1}^{k} m_h^{-1}\sum_{i=1}^{m_h} t_{jabcd}(\underline{\theta}^{**}_{cdhi},\underline{X}_{hi})\ e_{hia}\ .$$

Considering $\partial z_{2jabm}(\underline{\theta}_o)/\partial\theta_c$ first, since the derivatives of $\underline{\pi}$ and the elements of $V_{hi}(\underline{\theta},\underline{X})$ are uniformly bounded,

$$k^{-1}\sum_{h=1}^{k} m_h^{-1}\sum_{i=1}^{m_h} p_{jabc}(\underline{\theta}_o,\underline{X}_{hi})^2\ V(e_{hia})$$

converges to a constant depending on j, a, b, c and $\underline{\theta}_o$. Hence,

$$E\{\ [k^{-1}\sum_{h=1}^{k} m_h^{-1}\sum_{i=1}^{m_h} p_{jabc}(\underline{\theta}_o,\underline{X}_{hi})\ e_{hia}]^2\ \}$$

$$\le m^{-1}\ k^{-1}\sum_{h=1}^{k}(m/m_h)^{-1}\sum_{i=1}^{m_h} p^2_{jabc}(\underline{\theta}_o,\underline{X}_{hi})\ V(e_{hia})$$

$$= O(m^{-1})\ .$$

So by Corollary 5.1.1.1 in Fuller (1976),

$$k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} p_{jabc}(\underline{\theta}_o, \underline{X}_{hi})\ e_{hia} = O_p(m^{-1/2}) ,$$

implying that

$$\sum_{c=1}^{s+u} \partial z_{2jabm}(\underline{\theta}_o)/\partial \gamma_c\ (\hat{\theta}_c^{(0)} - \theta_{oc}) = O_p(m^{-1/2} a_m) .$$

For $\partial^2 z_{2jabm}(\underline{\theta}^{**})/\partial\gamma_c\partial\gamma_d$, Theorem 4.2 implies that

$$\lim_{k\to\infty} \lim_{m\to\infty} k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} t_{jabcd}(\underline{\theta}, \underline{X}_{hi})\ e_{hia} = 0 \quad \text{a.s.}$$

uniformly for all $\underline{\theta} \in \Phi \times \Gamma$. Thus,

$$k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} t_{jabcd}(\underline{\theta}_{cdhi}^{**}, \underline{X}_{hi})\ e_{hia} = o_p(1) ,$$

which implies that this is expression is $O_p(1)$. Hence

$$\sum_{c=1}^{s+u} \sum_{d=1}^{s+u} k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} \{ t_{jabcd}(\underline{\theta}_{cdhi}^{**}, \underline{X}_{hi})\ e_{hia}$$

$$\times (\hat{\theta}_c^{(0)} - \theta_{oc})\ (\hat{\theta}_d^{(0)} - \theta_{od}) \} = O_p(a_m^2) ,$$

and term 2 is $O_p[\max(m^{-1/2} a_m,\ a_m^2)]$.

Putting together the asymptotic expressions for term 1 and term 2 of $\hat{\underline{\gamma}}_m^{(1)} - \underline{\gamma}_o$,

$$\hat{\underline{\gamma}}_m^{(1)} - \underline{\gamma}_o = W_m^{-1}(\underline{\theta}_o)\ k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} \{ D(\underline{\gamma}_o, \underline{X}_{hi})'\ V^{-1}(\underline{\theta}_o, \underline{X}_{hi})$$

$$\times\ \underline{e}_{hi} \} + O_p[\max(m^{-1/2} a_m,\ a_m^2)] . \qquad \square$$

Note that the asymptotic expression for $\hat{\gamma}_m^{(1)} - \gamma_o$ in Theorem 4.4 is equal to

$$(\ddot{\gamma}_m - \gamma_o) + o_p[\max(m^{-1/2}a_m,\ a_m^2)] \ .$$

Since by Theorem 4.3 $m^{1/2}(\ddot{\gamma}_m - \gamma_o)$ converges in distribution,

$$(\ddot{\gamma}_m - \gamma_o) = O_p(m^{-1/2}) \ .$$

Furthermore, since $a_m < 1$,

$$\max(m^{-1/2}a_m,\ a_m^2) \le a_m \ .$$

Hence

$$(\ddot{\gamma}_m - \gamma_o) + o_p[\max(m^{-1/2}a_m,\ a_m^2)]$$

$$\le O_p[\max(m^{-1/2},\ a_m)] \ ,$$

which is less than or equal to $a_m$ if $m^{-1/2} \le a_m$. This implies that if the order in probability of the error in the initial estimator $\hat{\underline{\theta}}^{(0)}$ is greater than or equal to $m^{-1/2}$ (i.e., the error in the initial estimator is bounded in probability at a rate equal to or slower than $m^{-1/2}$), then the order in probability of error in $\hat{\gamma}_m^{(1)}$ is less than or equal to that of $\hat{\gamma}_m^{(0)}$.

Theorem 4.4 will now be used to prove the asymptotic normality of $\hat{\gamma}_m^{(1)}$.

*Theorem 4.5.* Assume that model (4.1) holds, that A4.1 - A4.8 and A4.10 - A4.12 hold, and let $\hat{\underline{\theta}}^{(0)}$ be a consistent estimator of $\underline{\theta}_o$ such that

$$\hat{\underline{\theta}}^{(0)} = \underline{\theta}_o + O_p(a_m)$$

for some sequence $\{a_m\}$ with $\lim_{m\to\infty} a_m = 0$ . Let

$$\underline{e}_{hi} = \underline{Y}_{hi} - \underline{\pi}(\underline{\gamma}_o, \underline{X}_{hi})$$

follow model (4.3) and let $\hat{\underline{\gamma}}_m^{(1)}$ be the one-step Gauss-Newton estimator defined in equation (4.7) with c = 1. Then

$$m^{1/2}(\hat{\underline{\gamma}}_m^{(1)} - \underline{\gamma}_o) \xrightarrow{L} N_s(\underline{0}, [k\, W(\underline{\theta}_o)]^{-1})$$

as $m\to\infty$, where

$$W(\underline{\theta}_o) = \lim_{m\to\infty} k^{-1} \sum_{h=1}^{k} m_h^{-1} \sum_{i=1}^{m_h} \{D(\underline{\gamma}_o, \underline{X}_{hi})' \, V^{-1}(\underline{\theta}_o, \underline{X}_{hi}) \times D(\underline{\gamma}_o, \underline{X}_{hi})\} .$$

*Proof.* By equation (4.5) and Theorem 4.4,

$$m^{1/2}(\hat{\underline{\gamma}}_m^{(1)} - \underline{\gamma}_o) = m^{1/2}(\ddot{\underline{\gamma}}_k - \underline{\gamma}_o) + O_p[\max\{a_m, m^{1/2}a_m^2\}] .$$

So by Theorem 4.3 and by Corollary 5.2.6.1 in Fuller (1976)

$$m^{1/2}(\hat{\underline{\gamma}}_m^{(1)} - \underline{\gamma}_o) \xrightarrow{L} N_s(\underline{0}, [k\, W(\underline{\theta}_o)]^{-1})$$

as $m\to\infty$. □

Note that since the order in probability of the error for $m^{1/2}(\hat{\underline{\gamma}}_m^{(1)} - \underline{\gamma}_o)$ must be less than one, we must have

$a_m < m^{-1/4}$. Hence to achieve an improved asymptotically normal one-step Gauss-Newton estimator, the order in probability for $\hat{\underline{\theta}}^{(0)}$ must lie in the interval $[m^{-1/2}, m^{-1/4})$.

#### 4.4.6. The c-step Gauss-Newton Estimator

The c-step Gauss-Newton estimator is defined in (4.7) and constructed iteratively starting with the one-step Gauss-Newton estimator. If $m^{-1/2} \le a_m < m^{-1/4}$, then by Theorem 4.5, $\hat{\underline{\gamma}}_m^{(c)}$ will have the same limiting distribution as $\hat{\underline{\gamma}}_m^{(1)}$. To see this, consider c = 2 and note that

$$\hat{\underline{\gamma}}_m^{(1)} = \underline{\gamma}_o + O_p(b_m) \; ,$$

where

$$b_m = \max\{a_m, m^{1/2} a_m^2\} \; .$$

Hence the order in probability of $m^{1/2}(\hat{\underline{\gamma}}_m^{(2)} - \underline{\gamma}_o)$ is

$$\max\{b_m, m^{1/2} b_m^2\} = \max\{a_m, m^{1/2} a_m^2\}.$$

# 5. ESTIMATORS OF CORRELATION BETWEEN BINARY RESPONSES FOR FAILURE TIME DATA COLLECTED FROM INDEPENDENT GROUPS OF CORRELATED INDIVIDUALS

## 5.1. Introduction

The previous chapters have described methods of estimating failure time distributions for data collected from independent groups of correlated individuals. The analyses involve estimating parameters in models for means of binary response vectors for each group and inspection interval. The elements of the mean vectors are hazard probabilities and are thus functions of the failure time distribution parameters. The covariance matrix for each binary response vector depends on the mean vector and parameters describing the correlations among the elements of the observed response vector. Multivariate nonlinear least squares estimation based on a Gauss-Newton algorithm is used to obtain estimates of the failure time distribution parameters. When the Gauss-Newton iterations are initiated with consistent estimates of the mean model and correlation parameters, the estimators have a joint asymptotic normal distribution. Many techniques exist for obtaining consistent estimates of the mean model parameters, but little information is available on consistent estimation of correlation coefficients for clustered binary data.

In this chapter, estimators of the correlation coefficient for the conditional binary variables are considered. Sufficient conditions for the consistency of these estimators are presented, and results from a small simulation to evaluate the performance of these estimators are discussed. The sufficiency conditions and simulation results are used to assess the advantages and disadvantages for each estimator under various model and data conditions.

## 5.2. Correlation Estimators

This section is concerned with developing consistent estimators of the correlation between binary responses for individuals belonging to the same group who are at risk during a specific interval. For data collected from groups of individuals, consistency implies that the estimator converges in probability to the true correlation as the number of groups increases. Hence, pairs of individuals whose responses have a common true correlation must be distributed across groups in order for a consistent estimator of the true correlation coefficient to be developed. For the purposes of this chapter, the correlation between binary responses for individuals belonging to the same group will be assumed to be homogeneous within and across groups for a particular interval. Although this assumption appears to be restrictive, results based on this condition can be applied to groups with heterogeneous correlation structures for which pairs of individuals can be categorized into correlation classes (e.g., male-male, male-female, female-female). In this case, the results are applied to each correlation class.

To construct a consistent estimator of the correlation, estimators for the covariances and variances of the conditional binary variables are needed, and the estimator must pool information over groups in some fashion. To understand how the proposed consistent correlation estimators are constructed, estimation of the intra-group correlation for one group is initially considered. The averaging process for the development of consistent correlation estimators is then discussed.

The data for estimation consist of conditional binary responses, $Y_{hij}$, for individual j in group i during interval h, where

$Y_{hij} = 1$ if individual j in group i fails during interval h given success in the previous interval,

$\quad\;\; = 0$ if individual j in group i succeeds during interval h given success in the previous interval,

$j = 1, 2, \ldots, n_{hi}$ individuals in group i during interval h, $i = 1, 2, \ldots, m_h$ groups in interval h, and $h = 1, 2, \ldots, k$ intervals. $Y_{hij}$ remains undefined if the individual has previously failed or is censored during or prior to interval h. Let $\pi_{hij}$ denote the mean of $Y_{hij}$, and recall that $\pi_{hij}$ is the hazard probability for individual j in group i during interval h. The variance of $Y_{hij}$ is $\pi_{hij}(1-\pi_{hij})$. Let $\hat{\pi}_{hij}$ be a consistent estimator of the mean as the number of groups, m, gets large.

Consider estimation of $\rho_{hi}$, the homogeneous intra-group correlation between the elements of a single response vector for group i during interval h. One possible estimator for $\rho_{hi}$ is

$$\hat{\rho}_{1hi} = \frac{\sum\sum_{j>j'} (Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'})}{\sum\sum_{j>j'} \left[(Y_{hij}-\hat{\pi}_{hij})^2 (Y_{hij'}-\hat{\pi}_{hij'})^2\right]^{1/2}}$$

$$= \frac{\sum_{j>j'}\sum (Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'})}{\sum_{j>j'}\sum |(Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'})|} .$$

Note that $\hat{\rho}_{1hi}$ will always provide an estimate that lies in the parameter space of the true correlation $\rho_{hi}$. However, when pooling over groups, estimators based on $\hat{\rho}_{1hi}$ are not consistent, and hence estimators which pool $\hat{\rho}_{1hi}$ over groups are not considered further. An alternative estimator, based on $\hat{\pi}_{hij}(1-\hat{\pi}_{hij})$ rather than $(Y_{hij}-\hat{\pi}_{hij})^2$ as an estimator for $\mathrm{Var}(Y_{hij})$, is

$$\hat{\rho}_{2hi} = \frac{\sum_{j>j'}\sum (Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'})}{\sum_{j>j'}\sum \left[\hat{\pi}_{hij}(1-\hat{\pi}_{hij})\hat{\pi}_{hij'}(1-\hat{\pi}_{hij'})^2\right]^{1/2}} . \qquad (5.1)$$

While $\hat{\pi}_{hij}(1-\hat{\pi}_{hij})$ provides a better estimate of $\mathrm{Var}(Y_{hij})$ than $(Y_{hij}-\hat{\pi}_{hij})^2$, $\hat{\rho}_{2hi}$ has the disadvantage of not being constrained to the parameter space. For example, if $n_{hi}$ = 2, $\hat{\pi}_{hi1} = \hat{\pi}_{hi2} = .4$, $Y_{hi1} = Y_{hi2} = 1$, then $\hat{\rho}_{2hi} = 1.5$.

Another possible pair of estimators for $\rho_{hi}$ can be fashioned after an estimator cited in Morrison (1976) for the case when individual variances within groups are homogeneous. The first estimator, based on $(Y_{hij}-\hat{\pi}_{hij})^2$ as an estimate of $\mathrm{Var}(Y_{hij})$, is

$$\hat{\rho}_{3hi} = \frac{\sum\limits_{j>j'}\sum\limits_{j'} (Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'}) \;/\; [n_{hi}(n_{hi}-1)/2]}{\left[ \sum\limits_{j} (Y_{hij}-\hat{\pi}_{hij})^2/n_{hi} \;\; \sum\limits_{j'} (Y_{hij'}-\hat{\pi}_{hij'})^2/n_{hi}\right]^{1/2}}$$

$$= \frac{\sum\limits_{j>j'}\sum\limits_{j'} (Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'}) \;/\; [n_{hi}(n_{hi}-1)/2]}{\sum\limits_{j} (Y_{hij}-\hat{\pi}_{hij})^2 \;/\; n_{hi}} , \tag{5.2}$$

where $n_{hi}$ is the number of individuals in group i who succeeded in interval h-1 and are not censored during interval h. Note that

$$\hat{\rho}_{3hi} = \frac{\sum\limits_{j}\sum\limits_{j'} (Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'}) - \sum\limits_{j} (Y_{hij}-\hat{\pi}_{hij})^2}{(n_{hi}-1) \;\; \sum\limits_{j} (Y_{hij}-\hat{\pi}_{hij})^2}$$

$$= \frac{1}{(n_{hi}-1)} \left\{ \frac{\left[ \sum\limits_{j} (Y_{hij}-\hat{\pi}_{hij})\right]^2}{\sum\limits_{j} (Y_{hij}-\hat{\pi}_{hij})^2} - 1 \right\} .$$

Since

$$0 \le \left[ \sum_{j} (Y_{hij}-\hat{\pi}_{hij})\right]^2 \le \sum_{j} (Y_{hij}-\hat{\pi}_{hij})^2 ,$$

the bounds on $\hat{\rho}_{3hi}$ are

$$-(n_{hi}-1)^{-1} \le \hat{\rho}_{3hi} \le 1,$$

indicating that the possible values for this estimator form a subset of the correlation parameter space. The lower bound of $-(n_{hi}-1)^{-1}$ will not in general be restrictive since intra-group correlations among individuals are typically positive. Further, a sufficient condition for the consistency of estimators based on (5.2) [and (5.3) below] is that correlations and variances are constant for a given group. In this case, the parameter space for the true correlation $\rho_{hi}$ is identical to the space defined by the bounds on $\hat{\rho}_{3hi}$.

By estimating $\mathrm{Var}(Y_{hij})$ with $\hat{\pi}_{hij}(1-\hat{\pi}_{hij})$, an alternative estimator for the correlation between two responses in group i during interval h is

$$\hat{\rho}_{4hi} = \frac{\sum\sum_{j>j'} (Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'}) \;/\; [n_{hi}(n_{hi}-1)/2]}{\sum_j \hat{\pi}_{hij}(1-\hat{\pi}_{hij}) \;/\; n_{hi}} \,. \tag{5.3}$$

As with $\hat{\rho}_{2hi}$, it is possible for values of this estimator to assume values larger than one or smaller than minus one. For example, if $n_{hi} = 2$, $\hat{\pi}_{hij} \equiv .4$, $Y_{hi1} = Y_{hi2} = 1$, then $\hat{\rho}_{2hi} = 1.5$.

Consider the case where intra-group responses during a specific interval h for a sample of groups have an intra-group correlation that is constant across all groups. Denote the common true value of the correlation $\rho_h$. Pooling information across groups to estimate $\rho_h$ can be achieved via ratio estimation, in which averaging is performed separately for the numerator and denominator of the ratios in (5.1) - (5.3), or via averaging of the ratios. Ratio estimators for $\rho_h$ based on (5.1) - (5.3) for the case when intra-group correlations are assumed to be homogeneous within and across groups for interval h are constructed as follows:

$$\hat{\rho}_{2h} = \frac{\sum_i \sum_{j>j'} \sum (Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'})}{\sum_i \sum_{j>j'} \sum \left[\hat{\pi}_{hij}(1-\hat{\pi}_{hij})\hat{\pi}_{hij'}(1-\hat{\pi}_{hij'})^2\right]^{1/2}} \tag{5.4}$$

$$\hat{\rho}_{3h} = \frac{\sum_i \sum_{j>j'} \sum (Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'}) \;/\; [n_{hi}(n_{hi}-1)/2]}{\sum_i \sum_j (Y_{hij}-\hat{\pi}_{hij})^2 \;/\; n_{hi}} \tag{5.5}$$

$$\hat{\rho}_{4h} = \frac{\sum_i \sum_{j>j'} \sum (Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'}) \;/\; [n_{hi}(n_{hi}-1)/2]}{\sum_i \sum_j \hat{\pi}_{hij}(1-\hat{\pi}_{hij}) \;/\; n_{hi}} \,. \tag{5.6}$$

In Section 5.3, estimator (5.4) will be shown to be consistent under quite broad conditions. However, sufficient conditions for the consistency of estimators (5.5) and (5.6) include the more restrictive condition of homogeneous variances within groups, although variances may vary across intervals and groups. This condition is met, for example, when all members of a group have the same marginal survival distribution.

Estimators for which the ratios in (5.1) - (5.3) are averaged over groups can also be constructed, although ratio estimators are generally less variable for this type of problem. These averages were considered in the simulation described in Section 5.3, but as predicted, their performance was poor and they will not be addressed in this chapter.

If the common intra-group correlation is assumed to be constant across intervals as well as groups, the ratio estimators in (5.4) - (5.6) can be pooled across intervals to obtain consistent estimators of the correlation. Using ratio estimation techniques, estimators of the common correlation $\rho$ are defined to be

$$\hat{\rho}_2 = \frac{\sum_h m_h^{-1} \sum_i \sum_{j>j'} (Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'})}{\sum_h m_h^{-1} \sum_i \sum_{j>j'} \left[\hat{\pi}_{hij}(1-\hat{\pi}_{hij})\hat{\pi}_{hij'}(1-\hat{\pi}_{hij'})\right]^{1/2}} \tag{5.7}$$

$$\hat{\rho}_3 = \Big[ \sum_h m_h^{-1} \sum_i \sum_j (Y_{hij}-\hat{\pi}_{hij})^2 \; / \; n_{hi} \Big]^{-1}$$

$$\times \Big[ \sum_h m_h^{-1} \sum_i \sum_{j>j'}\sum (Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'})$$

$$/ \; [n_{hi}(n_{hi}-1)/2] \Big] \tag{5.8}$$

$$\hat{\rho}_4 = \Big[ \sum_h m_h^{-1} \sum_i \sum_j \hat{\pi}_{hij}(1-\hat{\pi}_{hij}) \; / \; n_{hi} \Big]^{-1}$$

$$\times \Big[ \sum_h m_h^{-1} \sum_i \sum_{j>j'}\sum (Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'})$$

$$/ \; [n_{hi}(n_{hi}-1)/2] \Big] \; . \tag{5.9}$$

## 5.3. Consistency of Correlation Estimators

### 5.3.1. Assumptions

One condition that is used to show that the ratio estimators converge in probability to $\rho_h$ as the number of groups, m, gets large is that the denominator of each estimator converges in probability to a positive number. Assumption A5.1 is applied to estimator (5.4), and assumption A5.2 is applied to estimators (5.5) and (5.6).

*Assumption A5.1.* For some $c > 0$,

$$m_h^{-1} \sum_i \sum_{j>j'} \sum \left[\pi_{hij}(1-\pi_{hij})\pi_{hij'}(1-\pi_{hij'})\right]^{1/2}$$

converges in probability to $c$ as $m \to \infty$.

*Assumption A5.2.* For some $c > 0$,

$$m_h^{-1} \sum_i [n_{hi}(n_{hi}-1)/2]\ \pi_{hij}(1-\pi_{hij})$$

converges in probability to $c$ as $m \to \infty$.

As in Chapter 4, an assumption about switching limits from $m$ to $m_h$ is needed to demonstrate convergence in probability as $m$, the number of groups in the sample, gets large. This assumption is required for convergence of the numerators of the estimators.

*Assumption A5.3.* For any $h$,

$$\lim_{m\to\infty} m_h^{-1} \sum_i \sum_{j>j'} \sum (Y_{hij}-\pi_{hij})(Y_{hij'}-\pi_{hij'})$$

$$= \lim_{m_h\to\infty} m_h^{-1} \sum_i \sum_{j>j'} \sum (Y_{hij}-\pi_{hij})(Y_{hij'}-\pi_{hij'})$$

whenever the right hand side limit exists.

*Assumption A5.4.* For any $h$,

$$\lim_{m\to\infty} m_h^{-1} \sum_i \sum_{j>j'} \sum (Y_{hij}-\hat{\pi}_{hi})(Y_{hij'}-\hat{\pi}_{hi}) \;/\; [n_{hi}(n_{hi}-1)/2]$$

$$= \lim_{m_h \to \infty} m_h^{-1} \sum_i \sum_{j>j'} \sum \{ (Y_{hij}-\hat{\pi}_{hi})(Y_{hij'}-\hat{\pi}_{hi}) / [n_{hi}(n_{hi}-1)/2] \}$$

whenever the right hand side limit exists.

A finite maximum group size is also assumed to exist.

*Assumption A5.5.* There exists a finite maximum group size for the population of groups, denoted by n.

In the context of data collected from independent groups, group sizes are not expected to increase without bound, making such an assumption reasonable.

One additional assumption is used to prove the consistency of estimators (5.5) and (5.6). While estimators (5.5) and (5.6) can be shown to be consistent under the following assumption, they may not provide consistent estimates in more general situations.

*Assumption A5.6.* For each h and i, $\pi_{hij}$ is constant for all j; i.e., $\pi_{hij} = \pi_{hi}$ for all j.

Consistency for all estimators can be demonstrated by showing that as the number of groups gets large, the numerator and denominator divided by the number of groups converge in probability to means of covariances and means of products of standard deviations, respectively. The proof for the consistency of estimator (5.4) is described in detail below. Proofs for the consistency of the other two estimators rely on the same argument, and hence will only be briefly outlined. The consistency of estimators (5.7) - (5.9) follow immediately from that of (5.4) - (5.6), respectively. The notation $o_p(1)$ refers to convergence in probability as the number of groups, m, gets large.

### 5.3.2. Consistency of $\hat{\rho}_{2hi}$

*Theorem 5.1.* Suppose that assumptions A5.1, A5.3 and A5.5 hold, and that the binary variable $Y_{hij}$ has mean $\pi_{hij}$ for j = 1, 2, ..., $n_{hi}$ individuals in group i, i = 1, 2, ..., $m_h$, during interval h, h = 1, 2, ..., k. Consider any interval, say, interval h. Let $\rho_h \equiv \mathrm{Corr}(Y_{hij}, Y_{hij'})$ for all i and j ≠ j′, and let m denote the number of groups present in the sample. Let $\hat{\pi}_{hij}$ be a consistent estimator such that for each h, i, and j,

$$\lim_{m\to\infty} \hat{\pi}_{hij} = \pi_{hij} \text{ in probability.}$$

Then

$$\hat{\rho}_{2h} = \frac{\sum\limits_{i} \sum\limits_{j>j'}\sum (Y_{hij}-\hat{\pi}_{hij})(Y_{hij'}-\hat{\pi}_{hij'})}{\sum\limits_{i} \sum\limits_{j>j'}\sum \left[\hat{\pi}_{hij}(1-\hat{\pi}_{hij})\hat{\pi}_{hij'}(1-\hat{\pi}_{hij'})\right]^{1/2}}$$

converges in probability to $\rho_h$ as $m\to\infty$.

*Proof.* Pick any h. The numerator of $\hat{\rho}_{2h}$ divided by $m_h$ can be expressed as

$$\begin{aligned} & m_h^{-1} \sum_{i} \sum_{j>j'}\sum (Y_{hij}-\pi_{hij})(Y_{hij'}-\pi_{hij'}) \\ & \quad + m_h^{-1} \sum_{i} \sum_{j>j'}\sum (Y_{hij}-\pi_{hij})(\pi_{hij'}-\hat{\pi}_{hij'}) \\ & \quad + m_h^{-1} \sum_{i} \sum_{j>j'}\sum (Y_{hij}-\hat{\pi}_{hij})(\pi_{hij'}-\pi_{hij'}) \\ & \quad + m_h^{-1} \sum_{i} \sum_{j>j'}\sum (\pi_{hij}-\hat{\pi}_{hij})(\pi_{hij'}-\hat{\pi}_{hij'}) \\ & = \text{term 1} + \text{term 2} + \text{term 3} + \text{term 4}\ . \end{aligned}$$

Since

$$\begin{aligned} & E\{(Y_{hij}-\pi_{hij})(Y_{hij'}-\pi_{hij'})\} \\ & \quad = \rho_h\ [\pi_{hij}(1-\pi_{hij})\pi_{hij'}(1-\pi_{hij'})]^{1/2}\ , \end{aligned}$$

it follows from Theorem 5.1.1. in Chung (1974) and assumption A5.5 that term 1 converges in probability as $m\to\infty$ to

$$m_h^{-1} \sum_{i} \sum_{j>j'}\sum \rho_h\ \pi_{hij}(1-\pi_{hij})\ \pi_{hij'}(1-\pi_{hij'})$$

if

$$\mathrm{Var}\{\sum_{j>j'}\sum (Y_{hij}-\pi_{hij})(Y_{hij'}-\pi_{hij'})\}$$

has a common bound for all i. Since

$$|(Y_{hij}-\pi_{hij})(Y_{hij'}-\pi_{hij'})| \le 1 ,$$

it follows that

$$|E\{(Y_{hij}-\pi_{hij})(Y_{hij'}-\pi_{hij'})\}| \le 1$$

and

$$|(Y_{hij}-\pi_{hij})(Y_{hij'}-\pi_{hij'}) -$$

$$E\{(Y_{hij}-\pi_{hij})(Y_{hij'}-\pi_{hij'})\}| \le 2 .$$

Hence by assumptions A5.5

$$\mathrm{Var}\{\sum_{j>j'}\sum (Y_{hij}-\pi_{hij})(Y_{hij'}-\pi_{hij'})\}$$

$$= \sum_{j>j'}\sum \sum_{l>l'}\sum E\Big\{[(Y_{hij}-\pi_{hij})(Y_{hij'}-\pi_{hij'})$$

$$- E\{(Y_{hij}-\pi_{hij})(Y_{hij'}-\pi_{hij'})\}]$$

$$\times [(Y_{hil}-\pi_{hil})(Y_{hil'}-\pi_{hil'})$$

$$- E\{(Y_{hil}-\pi_{hil})(Y_{hil'}-\pi_{hil'})\}]\Big\}$$

$$\le 4n(n-1) .$$

Thus, for term 1,

$$m_h^{-1} \sum_i \sum_{j>j'}\sum (Y_{hij}-\pi_{hij})(Y_{hij'}-\pi_{hij'})$$

$$= \rho_h \; m_h^{-1} \sum_i \sum_{j>j'} \sum [\pi_{hij}(1-\pi_{hij})\pi_{hij'}(1-\pi_{hij'})]^{1/2}$$

$$+ o_p(1).$$

For terms 2 and 3, it follows from

$$|(Y_{hij}-\pi_{hij})| \le 1$$

and

$$\hat{\pi}_{hij} - \pi_{hij} = o_p(1)$$

that

$$(Y_{hij}-\pi_{hij})(\pi_{hij'} - \hat{\pi}_{hij'}) = o_p(1) \;,$$

implying that terms 2 and 3 are $o_p(1)$. Finally, term 4 is $o_p(1)$ since $\hat{\pi}_{hij}$ is consistent. Hence, the numerator of $\hat{\rho}_{2h}$ is

$$\rho_h \; m_h^{-1} \sum_i \sum_{j>j'} \sum [\pi_{hij}(1-\pi_{hij})\pi_{hij'}(1-\pi_{hij'})]^{1/2} + o_p(1) \;.$$

The denominator of $\hat{\rho}_{2h}$ divided by $m_h$ is a continuous function of $\{\hat{\pi}_{hij}\}$, and thus can be expressed as

$$m_h^{-1} \sum_i \sum_{j>j'} \sum [\pi_{hij}(1-\pi_{hij})\pi_{hij'}(1-\pi_{hij'})]^{1/2} + o_p(1) \;.$$

Finally, by assumption A5.1,

$$m_h^{-1} \sum_i \sum_{j>j'} \sum [\pi_{hij}(1-\pi_{hij})\pi_{hij'}(1-\pi_{hij'})]^{1/2} > 0 \;.$$

Since $\hat{\rho}_{2h}$ is a continuous function of the numerator and denominator, $\hat{\rho}_{2h}$ converges in probability to $\rho_h$. □

### 5.3.3. Consistency of $\hat{\rho}_{3hi}$

*Theorem 5.2.* Suppose that assumptions A5.2 and A5.4 - A5.6 hold, and that the binary variable $Y_{hij}$ has mean $\pi_{hi}$ for j = 1, 2, ..., $n_{hi}$ individuals in group i, i = 1, 2, ..., $m_h$, during interval h, h = 1, 2, ..., k. Consider any interval, say, interval h. Let $\rho_h \equiv \mathrm{Corr}(Y_{hij}, Y_{hij'})$ for all i and j $\neq$ j', and let m denote the number of groups present in the sample. Let $\hat{\pi}_{hi}$ be a consistent estimator such that for each h and i,

$$\lim_{m\to\infty} \hat{\pi}_{hi} = \pi_{hi} \text{ in probability.}$$

Then

$$\hat{\rho}_{3h} = \frac{\sum_i \sum_{j>j'} \sum (Y_{hij}-\hat{\pi}_{hi})(Y_{hij'}-\hat{\pi}_{hi}) \;/\; [n_{hi}(n_{hi}-1)/2]}{\sum_i \sum_j (Y_{hij}-\hat{\pi}_{hi})^2 \;/\; n_{hi}}$$

converges to $\rho_h$ in probability as $m\to\infty$.

*Proof.* Pick any h. From the proof of Theorem 5.1,

$$(Y_{hij}-\hat{\pi}_{hi})(Y_{hij'}-\hat{\pi}_{hi'})$$
$$= (Y_{hij}-\pi_{hi})(Y_{hij'}-\pi_{hi'}) + o_p(1)$$

and

$$\mathrm{Var}\{ \sum_{j>j'}\sum (Y_{hij}-\hat{\pi}_{hi})(Y_{hij'}-\hat{\pi}_{hi'}) \;/\; [n_{hi}(n_{hi}-1)/2] \}$$

is bounded, so that the numerator of $\hat{\rho}_{3h}$ divided by $m_h$ can be written

$$\rho_h \, m_h^{-1} \sum_i \sum_{j>j'}\sum \pi_{hi}(1-\pi_{hi}) + o_p(1) \;.$$

Using the same argument for j=j′, the denominator divided by $m_h$ is equal to

$$m_h^{-1} \sum_i \sum_{j>j'}\sum \pi_{hi}(1-\pi_{hi}) + o_p(1) \;.$$

Hence $\hat{\rho}_{3h}$ converges to $\rho_h$ in in probability. □

### 5.3.4. Consistency of $\hat{\rho}_{4hi}$

*Theorem 5.3.* Suppose that assumptions A5.2 and A5.4 - A5.6 hold, and that the binary variable $Y_{hij}$ has mean $\pi_{hi}$ for $j = 1, 2, \ldots, n_{hi}$ individuals in group i, $i = 1, 2, \ldots, m_h$, during interval h, $h = 1, 2, \ldots, k$. Consider any interval, say, interval h. Let $\rho_h \equiv \mathrm{Corr}(Y_{hij}, Y_{hij'})$ for all i and $j \neq j'$, and let m denote the number of groups present in the sample. Let $\hat{\pi}_{hi}$ be a consistent estimator such that for each h and i,

$$\lim_{m\to\infty} \hat{\pi}_{hi} = \pi_{hi} \text{ in probability.}$$

Then

$$\hat{\rho}_{4h} = \frac{\sum_i \sum_{j>j'}\sum (Y_{hij}-\hat{\pi}_{hi})(Y_{hij'}-\hat{\pi}_{hi}) \;/\; [n_{hi}(n_{hi}-1)/2]}{\sum_i \hat{\pi}_{hi}(1-\hat{\pi}_{hi})}$$

converges in probability to $\rho_h$ as $m\to\infty$.

*Proof.* Pick any h. From the proof of Theorem 5.2, the numerator of $\hat{\rho}_{4h}$ divided by $m_h$ is equal to

$$\rho_h\, m_h^{-1} \sum_i \sum_{j>j'}\sum \pi_{hi}(1-\pi_{hi}) + o_p(1) \;.$$

Since the denominator divided by $m_h$ is a continuous function of $\{\hat{\pi}_{hi}\}$, it can be written as

$$m_h^{-1} \sum_i \sum_{j>j'}\sum \pi_{hi}(1-\pi_{hi}) + o_p(1) \;.$$

Hence $\hat{\rho}_{4h}$ converges to $\rho_h$ in in probability. □

## 5.4. Simulation to Evaluate Performance of Estimators

### 5.4.1. Simulation Objectives and Design

A simulation was planned to evaluate the performance of correlation estimators for conditional binary variables constructed from failure time data that are collected from independent groups of correlated individuals. The objective of the study was to investigate the effect of several factors on the bias of these estimators, and to check consistency by seeing if bias improves as the number of groups in the sample increases.

Factors under consideration included the value of the true correlation coefficient, the underlying failure time distribution, the number of groups, and the degree to which the generated survival data are interval censored. Because of the interest in determining how interval censoring affects the estimators, the common intra-group correlation was assumed to be constant across time as well as across groups. Hence estimators (5.7) - (5.9) were considered in the simulation.

Two levels for the true correlation, failure time distribution and interval censoring factors and three levels for the number of groups were selected. These levels are presented in Table 5.1. The two interval censoring schemes were selected to simulate interval censored data from a regular inspection schedule (10 intervals of length 5 units) and interval censored data that more closely approximate exact time data (50 intervals of length 1 unit). For the failure time distributions, shape parameters were chosen to provide specific distributional shapes (monotonically decreasing and unimodal densities). The scale parameters for the failure time distributions were determined such that for the given shape parameter, 99% of the failures occur by $t = 50$. Note that since the failure time distribution is constant for all individuals (i.e., no explanatory variable

Table 5.1. Specific factors investigated in the simulation to assess the performance of correlation estimators

| Factor | Levels |
| --- | --- |
| True Correlation | 0.3<br>0.6 |
| Failure Time Distribution | Monotonically decreasing Weibull density (scale = 6.5, shape = 0.75)<br>Unimodal Weibull density (scale = 23, shape = 2) |
| Number of Groups | 20<br>50<br>100 |
| Interval Censoring Scheme | [0,10) [10,20) ... [40,50) [50,∞)<br>[0,1) [1,2) ... [49,50) [50,∞) |

effects were included), each response for a given interval has the same mean and variance. This restriction insures that the data generation scheme provides constant correlation among members of a group. It also provides suitable conditions for assessing the consistency of estimators (5.8) and (5.9).

Because of limited resources, the simulation was run only for small group sizes. The distribution used to randomly select the number of individuals in a given group was patterned after the rat litter data discussed in Chapter 3, which typically consisted of four individuals per group. The probabilities that a group contains two, three or four individuals were set at.005, .070 and .925, respectively.

Data generation took place in two phases. First, data were generated according to the following scheme for 50 intervals of length one for each individual in each of 50 or 100 groups according to a specified correlation and failure time distribution. For each group, a random group size was generated. Then binary responses were generated for each group member at risk during the interval. This was accomplished for each interval by generating a value for a Bernoulli($\pi_h$) variable, $Y_{h0}$, where $\pi_h$ is the hazard probability for interval h calculated from the assumed failure time distribution. This variable was used to correlate the responses of each member by generating the response of the j-th individual in group i during interval h in the following way. Let the true correlation be denoted $\rho$, and define $Y_{-1,ij} = 0$. Given $Y_{h-1,ij} = 0$,

$$
\begin{aligned}
Y_{hij} &= \text{the value of } Y_{h0} \text{ with probability } \rho^{1/2}, \\
&= \text{a new value generated from a Bernoulli}(\pi_h) \\
&\qquad \text{distribution with probability } 1 - \rho^{1/2}.
\end{aligned}
$$

No value was assigned for individuals that failed in a previous interval, and no individuals were right censored until 50 time units.

For each level of correlation and failure time distribution, the data generation algorithm resulted in sets of $Y_{hij}$ values that had a common intra-group correlation coefficient for all intervals. These data were arranged in a grid with rows corresponding to the individuals from all of the 50 or 100 groups, and columns corresponding to the 50 intervals. For the second phase of the program, the correlation estimates were calculated for each level of the number of groups and interval censoring scheme factors by subsetting the grid for the correct number of groups and collapsing responses over intervals for the interval censoring factor when necessary. One hundred replications of the entire algorithm were conducted.

The simulation was run in two phases. Initially, the maximum number of groups under consideration was 50, and grids were subsetted to obtain 20 groups. However, results from 100 replications of all combinations of correlation, failure time distribution, number of groups (20 or 50) and interval censoring scheme indicated that bias did not improve for most estimators when comparing the average bias for 20 and 50 groups. Hence another set of 100 runs was conducted for 100 groups and each combination of correlation, failure time distribution and interval censoring scheme to further check the consistency property of the estimators. This design confounds comparisons involving 20 and/or 50 groups versus 100 groups with the simulation experiment; however, the mean values are not greatly influenced by the confounding.

To analyze the simulation results, an ANOVA was run on the bias for each estimator with correlation and failure time distribution as "whole plot" factors arranged in a completely randomized design, and number of groups and interval censoring scheme as the "split plot" factors. F-values from the ANOVA were used to identify the most important factors affecting the bias of each estimator. Sample standard deviations for each factor combination were also calculated to assess the variability of the estimators.

### 5.4.2. Simulation Results

Results from the ANOVAs on bias for each estimator indicate that the average bias for each estimator is not very large, and that the dominant factor influencing bias is the interval censoring scheme. All estimators tend to slightly underestimate $\rho$ for the partition of 10 five-unit intervals (Table 5.2). For the 50-interval partition, on average, both $\hat{\rho}_2$ and $\hat{\rho}_4$ slightly overestimate $\rho$, while $\hat{\rho}_3$ appears to be unbiased. The ANOVAs and the means listed in Table 5.2 also indicate that the number of groups in the data set seems to have little effect on the bias of the estimators.

Average estimated standard deviations for all estimators are quite high (Table 5.3). Standard deviations are about 25% larger for data involving 20 groups relative to standard deviations based on 50 or 100 groups. The same trend is present for the monotonically decreasing versus the unimodal distribution. Mixed results for the correlation and interval censoring factors are related to the type of variance estimator used in the denominator. Variability is about 25% higher for $\rho$=.6 relative to $\rho$=.3 and for the 50-interval relative to the 10-interval partition when $\hat{\pi}_{hij}(1-\hat{\pi}_{hij})$ is used to estimate $\mathrm{Var}(Y_{hij})$. The opposite trends are seen for $\hat{\rho}_3$, which is based on $(Y_{hij}-\hat{\pi}_{hij})^2$ as an estimator of $\mathrm{Var}(Y_{hij})$. It is not clear whether any of these differences are actually significant.

Inspection of the simulation results for parameter estimates that exceed the parameter space for correlations

Table 5.2. Average bias and standard errors for number of intervals and number of groups for each estimator

| Estimator | Number of Intervals | Number of Groups 20 | 50 | 100 | Mean |
|---|---|---|---|---|---|
| $\hat{\rho}_2$ | 10 | -.033[a] | -.040 | -.035 | -.036[b] |
| | 50 | .034 | .008 | .026 | .023 |
| $\hat{\rho}_3$ | 10 | -.036 | -.048 | -.038 | -.041 |
| | 50 | .0001 | -.008 | .008 | .000 |
| $\hat{\rho}_4$ | 10 | -.025 | -.042 | -.033 | -.033 |
| | 50 | .047 | .016 | .036 | .033 |

[a]Each cell mean is based on 400 observations. Standard errors of cell means for $\hat{\rho}_2$, $\hat{\rho}_3$ and $\hat{\rho}_4$ are .0068, .0063 and .0086, respectively.

[b]Each marginal mean is based on 1200 observations. Standard errors of marginal means for $\hat{\rho}_2$, $\hat{\rho}_3$ and $\hat{\rho}_4$ are .0035, .0036 and .0049, respectively.

Table 5.3. Average standard deviations for estimated correlation values for each level of each factor and for the entire data set.

| | | Estimator | | |
|---|---|---|---|---|
| Factor | Level | $\hat{\rho}_2$ | $\hat{\rho}_3$ | $\hat{\rho}_4$ |
| Number of Groups | 20 | .158[a] | .161 | .216 |
| | 50 | .115 | .131 | .173 |
| | 100 | .126 | .135 | .177 |
| Number of Intervals | 10 | .119[b] | .162 | .177 |
| | 50 | .148 | .123 | .200 |
| Distribution | Decreasing | .147[b] | .153 | .213 |
| | Unimodal | .120 | .132 | .164 |
| Correlation | .3 | .121[b] | .153 | .182 |
| | .6 | .146 | .131 | .195 |
| Entire Data Set | | .133[c] | .142 | .189 |

[a]Cell means are based on 8 estimated standard deviations.

[b]Cell means are based on 12 estimated standard deviations.

[c]Grand means are based on 24 estimated standard deviations.

indicated that estimated values for $\hat{\rho}_2$ and $\hat{\rho}_4$ exceed one only when the true correlation coefficient is .6. Of the 1200 values for $\rho = .6$, 1.9% are greater than one for $\hat{\rho}_2$ and 4.4% are greater than one for $\hat{\rho}_4$. Of the estimated values exceeding one for $\rho=.6$, nearly all occur for the 50-interval censoring scheme, about half are associated with data based on only 20 groups, and about two thirds are linked with monotonically decreasing distributions.

Since most groups contained four individuals, an approximate (but overestimated) lower bound on $\rho$ is -1/3. No $\hat{\rho}_2$ values and only one $\hat{\rho}_3$ and $\hat{\rho}_4$ value fall below -1/3 ($\hat{\rho}_3 = -.41$, $\hat{\rho}_4 = -.40$ in the same replication for $\rho = .3$, 20 groups, and the monotonically decreasing failure time distribution).

### 5.4.3. Discussion

Estimator $\hat{\rho}_2$ is suited for estimation when few restrictions are present on the structure of the means. Obtaining estimates exceeding one is the major problem associated with $\hat{\rho}_2$. This is a problem particularly for high correlations and more finely partitioned interval censoring, although the probability of obtaining estimates outside of the correlation parameter space is not particularly high (<.05). The risk of obtaining such a value is also reduced for larger numbers of groups and unimodal distributions.

Use of estimators $\hat{\rho}_3$ and $\hat{\rho}_4$ is appropriate when the individuals within a given group have the same intra-group correlations and variances. Variability is always larger for $\hat{\rho}_4$ than for $\hat{\rho}_3$, and bias for $\hat{\rho}_4$ is often worse than that of $\hat{\rho}_3$. In addition, $\hat{\rho}_4$ is not restricted to the parameter space, and behaves much more poorly than $\hat{\rho}_2$ in this respect. Hence $\hat{\rho}_3$ appears to be a better estimator with respect to bias, variability and parameter space constraints. It is also appealing that bias for $\hat{\rho}_3$ diminishes as the interval censoring scheme provides more precise information on failure times. Although $\hat{\rho}_2$ could also be used in this situation, it is clear that $\hat{\rho}_3$ is a better estimator under these conditions.

Although none of the estimators exhibits an obvious improvement in bias as the number of groups increases from 20 to 100, the variability of all the estimators declines as the number of groups in the sample increases. Further, average bias for the estimators is not particularly large. All estimators exhibit a large degree of variability indicating that any particular estimated value may not be very accurate.

As with all simulations, there are several limitations inherent in the design of this simulation study. The performance of these estimators for large group sizes may be quite different, and this is not addressed by the simulation. In addition, the effects of right censoring before study termination or of different truncation points prior to 50 time units are not considered. The presence of heterogeneous means, even across groups, is also not studied. It is possible to adapt the data generation scheme to allow means to vary across groups while maintaining a constant correlation. The performance of estimators for correlations that change over intervals could also be investigated by outputting estimates for each interval separately.

# 6. SUMMARY

A method of estimating failure time distributions for data collected from independent groups of correlated individuals is developed. This method is appropriate for commonly interval censored data (e.g. when individuals are inspected at regular common intervals) or exact time data, including possibly right censored data. The technique improves upon previously published methods by allowing for large and variable group sizes, heterogeneous correlation structures, and the incorporation of explanatory variable information. Both parametric and nonparametric failure time models can be estimated, and correlations may be modeled as well.

The outcome for any individual at risk during a specific time interval is modeled as a conditional binary random variable indicating the failure or success of the individual given success in the preceding interval. For each time interval, a separate vector of binary responses is constructed for each group, consisting of the responses for individuals belonging to the group who are at risk at the beginning of the interval and not censored during the interval. The elements of the corresponding mean vector are hazard probabilities and are thus functions of the failure

time distribution parameters. The covariance matrix for each binary response vector is a function of the corresponding mean vector and parameters describing the correlations among the elements of the observed response vector. Multivariate nonlinear least squares estimation based on the Gauss-Newton algorithm is used to obtain estimates of the parameters of the failure time distribution. These estimators are shown to have a joint asymptotic normal distribution under mild regularity conditions when a Gauss-Newton iterations are initiated with consistent estimates of the mean model and correlation parameters.

The new methodology is applied to commonly interval censored data from a study comparing the effectiveness of three smoking cessation programs, and to an exact time data set assessing the developmental responses of rat pups to prenatal doses of methylmercuric chloride. Results from both analyses indicate that estimated standard errors for the estimates of parameters associated with variables whose values differ between groups are generally higher when the analysis accounts for the presence of correlation. This result is expected since failing to account for correlation among group members results in the underestimation of standard errors of the parameter estimates, and thus overstatement of the significance of tests involving these parameters. Estimated standard errors that account for the presence of correlation may be smaller than independence-based estimates when the parameter is associated with a within-group explanatory variable, such as sex in the rat litter data.

When comparing independence-based estimates with least squares parameter estimates, it appears that parameters are not equally sensitive to the effects of correlation. For both data sets, only a subset of the estimated parameters obtained from the independence-based and least squares estimation procedures are significantly different from the corresponding subset obtained using least squares estimation. For exact time data, this shift may be due in part to treating the data more accurately as interval censored.

Research is also presented on the properties and performance of estimators of correlation coefficients for clustered binary data. Several consistent estimators are developed and their empirical performance is evaluated in a small simulation. The estimators are only slightly biased, but can be quite variable. The degree of interval censoring generally affects the sign, but not necessarily the magnitude of the bias. Increasing the number of groups from 20 to 100 does not provide a substantial reduction in bias, although variability does decline.

# 7. LITERATURE CITED

Buelke-Sam, J., C. A. Kimmel, J. Adams, C. J. Nelson, C. V. Vorhees, D. C. Wright, V. St. Omer, B. A. Korol, R. E. Butcher, M. A. Geyer, J. F. Holson, C. L. Kutscher and M. J. Wayner. 1985. Collaborative Behavioral Teratology Study: Results. Behavioral Toxicology and Teratology 7:591-624.

Chung, K. L. 1974. A course in probability theory. 2nd edition. Academic Press, New York, New York.

Clayton, D. G. 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65:141-151.

Clayton, D. and J. Cuzick. 1985. Multivariate generalizations of the proportional hazards model. Journal of the Royal Statistical Society, Series A 148:82-117.

Cox, D. R. and D. Oakes. 1984. Analysis of survival data. Chapman and Hall, New York.

Crowder, M. 1989. A multivariate distribution with Weibull connections. Journal of the Royal Statistical Society, Series B 51:93-107.

Fuller, W. A. 1976. Introduction to statistical time series. John Wiley and Sons, New York, New York.

Gallant, A. R. 1987. Nonlinear statistical methods. John Wiley and Sons, New York, New York.

Holt, J. D. and R. L. Prentice. 1974. Survival analyses in twin studies and matched pair experiments. Biometrika 61:17-30.

Hougaard, P. 1986. A class of multivariate failure time distributions. Biometrika 73:671-678.

Huster, W. J., R. Brookmeyer and S. G. Self. 1989. Modeling paired survival data with covariates. Biometrics 45:145-156.

Jennrich, R. I. 1969. Asymptotic properties of non-linear least squares estimators. Annals of Mathematical Statistics 40:633-643.

Kalbfleisch, J. D. and R. L. Prentice. 1980. The statistical analysis of failure time data. Wiley, New York, NY.

Koehler, K. J. and P. G. McGovern. 1990. Application of the LFP survival model to smoking cessation data. Statistics in Medicine 9:409-421.

Lando, H. A. 1977. Successful treatment of smokers with a broad-spectrum behavioral approach. Journal of Consulting and Clinical Psychology 45:361-366.

Lando, H. A. and P. G. McGovern. 1982. Three-year data on a behavioral treatment for smoking: A follow-up note. Addictive Behaviors 7:177-181.

Lando, H. A., P. G. McGovern, F. X. Barrios and B. D. Etringer. 1990. Comparative evaluation of American Cancer Society and American Lung Association smoking cessation clinics. American Journal of Public Health 80:554-559.

Lawless, J. F. 1980. Statistical methods for lifetime data. John Wiley and Sons, New York.

Liang, K. and S. L. Zeger. 1986. Longitudinal data analysis using generalized linear models. Biometrika 73:13-22.

McCullagh, P. and J. A. Nelder. 1983. Generalized linear models. Chapman and Hall, New York.

Meeker, W. Q. 1983. GENMAX: a program for maximum likelihood estimation of nonstandard statistical models. Department of Statistics, Iowa State University, Ames, IA.

Meeker, W. Q. 1987. Limited failure population life tests: application to integrated circuit reliability. Technometrics 29:51-66.

Morel, J. G. 1987. Multivariate nonlinear models for vectors of proportions: a generalized least squares approach. Unpublished doctoral dissertation, Iowa State University.

Morrison, D. F. 1976. Multivariate statistical methods. 2nd edition. McGraw-Hill, New York.

Oakes, D. 1982. A model for association in bivariate survival data. Journal of the Royal Statistical Society, Series B 44:414-412.

Oakes, D. 1989. Bivariate Survival models induced by frailties. Journal of the American Statistical Association 84:487-493.

Rao, C. R. 1973. Linear statistical inference and its applications. 2nd edition. John Wiley and Sons, New York, New York.

Rudin, W. 1976. Principles of mathematical analysis. McGraw-Hill Inc., New York, New York.

SAS Institute Inc. 1985. SAS user's guide: Statistics, Version 5 edition. SAS Institute, Inc., Cary, NC.

Wei, L. J. and D. A. Amato. 1989. Regression analyses for highly stratified failure time observations. Paper presented at 1989 ENAR Biometrics Meetings, Lexington, KY.

Zeger, S. L. and K. Liang. 1986. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42:121-130.

# 8. ACKNOWLEDGEMENTS

I would like to thank my husband, Michael King, for his unwavering support in the pursuit of this degree, and for always encouraging me to pursue my interests, whatever they may be. I would also like to thank my family, particularly my parents, Margaret and William Nusser, for their support and for the many needed respites they have provided us through the years.

My major professor, Ken Koehler, deserves a great deal of credit for his excellent guidance over the past three years, especially concerning many of the ideas in this thesis and for teaching me more about how to think about research and statistics from an intuitive perspective. I am particularly grateful for his dedication in helping me finish this thesis before the birth of Michael's and my child.

The support and advice of my other committee members, Bill Clark, Noel Cressie, Paul Hinz and Bill Meeker, are also gratefully acknowledged. In particular, Bill Meeker's excellent teaching abilities deepened my interest in failure time methods. Bill Clark's dedication to his role on my committee and his eagerness to be actively involved in my research is greatly appreciated.

I would also like to thank those who contributed to my development outside the domain of this thesis. While I was under their supervision, Dave Cox and Paul Hinz taught me a great deal about consulting and analyzing data. Wayne Fuller has been a wonderful teacher and mentor, and has been especially generous in providing me with many professional opportunities that are often not available to graduate students. Much of my work with Wayne Fuller was supported by the Center for Agricultural and Rural Development; in particular, Helen Jensen's active support of my professional endeavors has been invaluable.

I would like to thank Alicia Carriquiry and Karen Jensen, who have always kept me informed of the various problems I would encounter in this process, and whose friendship I have enjoyed and still greatly value. I am grateful to Sahadeb Sarkar for many helpful discussions on the theoretical aspects of this work. The secretarial and emotional support of both Helen Nelson and Norma Elwick is also greatly appreciated. Finally, thanks go to the many other unnamed staff and friends who have made my stay at ISU so enjoyable.